NALYSIS

Edited by Christopher Kirwan

16

Volume 47

1987

ISSN 0003-2638

BASIL BLACKWELL · OXFORD

INDEX

		PAGE
APPIAH, ANTHONY	'If' again	1
BLACKBURN, WILLIAM K.	Davidson on force and convention	72
BODANSZKY, EVA	Parfit on selves and their interests	4
BRENNAN, ANDREW A.	Survival and importance	225
BROACKES, JUSTIN	Thoughts and definitions	95
BURKE, MICHAEL B.	Theodicy with a God of limited power: a reply to McGrath	
CAHN, STEVEN M.	A puzzle concerning divestiture	17
CLARK, MICHAEL	The truth about heaps	177
COLTERJOHN, JULIA *and* MACINTOSH, DUNCAN	Gerald Vision and indexicals	58
CUMMISKEY, DAVID	Desert and entitlement: a Rawlsian consequentialist account	15
DAVIES, DAVID	How not to outSmart the anti-realist	
DUDMAN, V. H.	Appiah on 'if'	74
EDITORIAL		65
EHRING, DOUGLAS	Non-simultaneous causation	28
EHRING, DOUGLAS	Survival and trivial facts	5
EHRING, DOUGLAS	Compound emphasis and causal relata	209
ELLIOT, ROBERT	Moral realism and the modal argument	133
ENGLEBRETSEN, GEORGE	Morris on identity	92
FORBES, GRAEME	Free and classical counterparts: response to Lewis	14
GALE, RICHARD M.	Parfit's arguments against partially relativized theories of rationality	230
GARRETT, B. J.	A further reply to Noonan	204
GILL, R. R. ROCKINGHAM	Frege on infinite axiom-systems	17
GILLETT, G. R.	The generality constraint and conscious thought	20
HELM, PAUL	On pan-critical irrationalism	24
HOCHBERG, HERBERT	Quantification, description, and identity	87
HOOKER, BRAD	Williams' argument against external reasons	4
HORNSBY, JENNIFER	Reply to Wreen	238
HUNT, G. M. K.	Determinism, predictability and chaos	129
KENNEDY, RALPH	Salmon versus Kripke on the *a priori*	158

(U)
An 13

ol. 47 No. 1 (New Series No. 213) January 1987

ANALYSIS

cue-H02957-16-P004709

Edited by

CHRISTOPHER KIRWAN

CONTENTS

w not to outSmart the anti-realist	DAVID DAVIES
s, constraints and trumps	PHILIP PETTIT
ert and entitlement: a Rawlsian consequentialist account	DAVID CUMMISKEY
e generality constraint and conscious thought	G. R. GILLETT
pan-critical irrationalism	PAUL HELM
multaneous causation	DOUGLAS EHRING
l sufficiency reconsidered	STIG ALSTRUP RASMUSSEN
ining contrastive facts	DAVID-HILLEL RUBEN
an argument for incompatibilism	DAVID WIDERKER
iams' argument against external reasons	BRAD HOOKER
counterexample to *modus ponens*	E. J. LOWE
fit on selves and their interests	EVA BODANSZKY
vival and trivial facts	DOUGLAS EHRING
heism or agnosticism	P. J. MCGRATH
eodicy with a God of limited power: a reply to McGrath	MICHAEL B. BURKE
rald Vision and indexicals	JULIA COLTERJOHN *and* DUNCAN MACINTOSH
re and now	PAUL SIMPSON
ley on causation and rationality	SCOTT STURGEON

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003-2638

KVANVIG, JONATHAN L.	On Lemke's defence of a causal basing requirement	162
LAMARQUE, PETER	The puzzle of the flash stockman: a reply to David Lewis	93
LEE, RICHARD	Williams, *ought*, and logical form	137
LEON, MARK	Solipsism regained	116
LIPTON, PETER	A real contrast	207
LOUI, R. P.	Nozick's acceptance rule and the lottery paradox	213
OWE, E. J.	Not a counterexample to *modus ponens*	44
LOWE, E. J.	Lewis on perdurance versus endurance	152
LOWE, E. J.	Reply to Noonan	201
MCGRATH, P. J.	Atheism or agnosticism	54
MCGRATH, P. J.	Children of a lesser God? A reply to Burke and Crisp	236
MACINTOSH, DUNCAN	(*see* COLTERJOHN, JULIA)	
MONTMARQUET, JAMES	A first-person asymmetry	167
NATHAN, N. M. L.	Evidential insatiability	110
NOONAN, HAROLD W.	Reply to Spinks on temporal parts	187
OLIN, DORIS	On an epistemic paradox	216
OPPY, GRAHAM	Williamson and the contingent *a priori*	188
OVER, D. E.	Assumptions and the supposed counterexamples to *modus ponens*	142
PATTON, THOMAS E.	On a persistent fallacy regarding the *de re*	65
PETTIT, PHILIP	Rights, constraints and trumps	8
PRIEST, GRAHAM	Tense, *tense* and TENSE	184
RASMUSSEN, STIG ALSTRUP	Causal sufficiency reconsidered	32
RASMUSSEN, STIG ALSTRUP	Sense, reference, and meaning-incommensurability	170
REA, GEORGE	How many minds has a tredecim?	100
REA, GEORGE	A variation on Hollis's paradox	218
RUBEN, DAVID-HILLEL	Explaining contrastive facts	35
SAYWARD, CHARLES	Prior's theory of truth	83
SCHWARTZ, STEPHEN P.	Intuitionism and sorites	179
SIMPSON, PAUL	Here and now	61
STURGEON, SCOTT	Foley on causation and rationality	62
TEICHMANN, ROGER	*Tu quoque*	199
TSOHATZIDIS, SAVAS L.	Deontic trouble in speech act botany	80

VISION, GERALD	Antiphon	124
WHITE, ALAN R.	Visualizing and imagining seeing	22[illegible]
WHITE, ALAN R.	'"By" and "By"'	239
WIDERKER, DAVID	On an argument for incompatibilism	3[illegible]
WILLIAMS, C. J. F.	Knowledge, belief and existence	103
WILLIAMSON, TIMOTHY	On knowledge of the unknowable	154
WREEN, MICHAEL	Two 'by' ways	120

BLACKWELL

The Nature of Time

Edited by RAYMOND FLOOD
and MICHAEL LOCKWOOD

In this impressive collection of original articles ten internationally-known scholars use the latest developments in thermodynamics, quantum theory and cosmology to explore and explain the nature of time, space and space-time.
200 pages, **£19.50** (0 631 14807 8)

The Languages of Logic

S. D. GUTTENPLAN

'[This book] brings out clearly and succinctly the relation of formal logic to natural languages and provides the necessary grounding for topics treated in the philosophy of logic and language. Issues about reference, names and descriptions arise naturally from the treatment of predicate logic. Many textbooks fail in this respect—it will fill a real gap.' David Hirschmann
344 pages, hardback **£25.00** (0 631 14624 5)
paperback **£7.95** (0 631 14625 3)

Fact, Science and Morality

Essays on A. J. Ayer's *Language, Truth and Logic*
Edited by GRAHAM MACDONALD and CRISPIN WRIGHT

These essays celebrate the fiftieth anniversary of the publication of A. J. Ayer's *Language, Truth and Logic*, a book which has been of major importance in spreading the philosophy of logical positivism to an English-speaking readership. Each essay discusses a topic raised by Ayer, assessing the contemporary philosophical relevance of the positivist enterprise.
*c.*320 pages, *c.***£25.00** (0 631 14555 9)

Needs, Values, Truth

Essays in the Philosophy of Value
DAVID WIGGINS

This collection comprises three new essays, and revised or corrected versions of several published essays, including the author's recent 'Claims of Need', his two redeployments of Aristotle's theory of practical reason, and an attempt to refurbish the incompatibilist view of human freedom 'Towards a Reasonable Libertarianism'.
Aristotelian Society Series
352 pages, *c.***£27.50** (0 631 14044 1)

Metaphor

DAVID E. COOPER

'Why has metaphor become such an important topic?', 'what is metaphorical meaning?', 'what is metaphorical truth?'. David E. Cooper argues that metaphor is best understood without reference to metaphorical meaning and truth, and that its sustaining function is the social one of 'cultivating intimacy'.
Aristotelian Society Series
288 pages, **£25.00** (0 631 14938 4)

Basil Blackwell

108 Cowley Road, Oxford OX4 1JF
Suite 1503, 432 Park Avenue South, New York NY 10016

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol. 64, No. 4; December 1986

EDITOR: BRIAN ELLIS

ARTICLES	
Jordan Howard Sobel	Notes on Decision Theory: Old Wine in New Bottles
Philip Pettit and Geoffrey Brennan	Restrictive Consequentialism
D.H.M. Brooks	Group Minds
Bredo C. Johnsen	The Inverted Spectrum
Rod Bertolet	Donnellan's Distinctions
Karen Green	Psychologism and Anti-Realism
DISCUSSION	
Elizabeth Ring	Goldman's Reliability Theory of Justified Belief
Don Mannison	Another Failure for Cognitive Relativism

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTION: Individuals A$17 (US$18 or Stg£10); Institutions A$34, (US$36 or Stg£18). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

The Journal of Philosophy

SUBSCRIPTIONS

Individuals	$20.00
Libraries and institutions	$25.00
Students, retired/unemployed philosophers	$15.00
Postage outside U.S.	$ 5.00

BACK VOLUMES

Published bi-monthly 1904–1976, monthly thereafter.

All back volumes and individual issues are available, also a Cumulative Fifty-year Index, 1904–1953 and a Ten-year Supplemental Index, 1954–1963. Back volumes 1976–present also available in microform.

Please inquire for price lists and shipping charges on back orders.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

ANALYSIS 47.1 JANUARY 1987

HOW NOT TO OUTSMART THE ANTI-REALIST

By DAVID DAVIES

I

HILARY PUTNAM [6] [7] has argued that proponents of what he terms 'metaphysical realism' are committed to a thesis – that even an 'Ideal Theory' might be false – that cannot be consistently and coherently maintained. In recent issues of ANALYSIS, the merits of Putnam's contention have been debated by J. J. C. Smart, Peter Smith, and Norman Melchert. Smart [10] offers an example which purportedly illustrates how we *could* quite consistently suppose that an Ideal Theory might be false. Smith [11] responds that the proposed example establishes, at best, the possible *incompleteness* of an Ideal Theory. Melchert [5] seems prepared to grant the force of Smith's response to Smart, but he offers, in support of the latter's contention that metaphysical realism is correct, a different example which is supposedly immune to Smith's criticisms. And, most recently, Smith [12] rejects Melchert's example as having no bearing on the tenability of metaphysical realism as understood by Putnam and Smart. My aim, in this discussion, is to rescue Putnam's original claim from the ministrations of friend and foes alike. I shall argue for the following conclusions: (a) Melchert's example fails to engage the issue between Putnam and Smart because it misconstrues the relevant notion of an 'Ideal Theory'; (b) Smith's response to Smart is not available to *Putnam*, since it would effectively undermine the very argument which the latter directs against metaphysical realism, but (c) once this point is recognized, it is apparent that Putnam has a simpler reply to Smart – in effect, Smart's example begs the question unless Putnam's argument fails on *independent* grounds.

II

Melchert takes his example from the realm of human actions intentionally construed. The 'stories' we tell concerning this realm differ from our physical theories, he maintains, in that the evidence to which the former are accountable may 'somehow get lost or distorted irretrievably'. Consequently 'the ideal story from the point of view of a community of historical inquirers may still be false . . . The best evidence available, or *ever to be available*, may confirm a falsehood. And metaphysical realism is correct'. Thus a person may make a remark, not intending it as an insult, under circumstances such that the rational conclusion for the linguistic community to draw, at the time of the event and in the face of any evidence that will be available in the future, is that the remark *was* intended as an insult. In this case, 'the ideal story (the one that makes the best total sense of all the available evidence) describes the incident as an insult', and the ideal story therefore contains a *false* assertion.

Melchert's conception of an 'ideal story', as exemplified in the preceding quotation, is *not*, however, an appropriate analogue, with respect to particular human actions, for the conception of an Ideal Theory as the latter is understood in Putnam's argument against metaphysical realism. What makes a theory 'Ideal', in Putnam's sense, is not that it is the best theory *given the evidence we would ever, in practice, have available*, but, rather, that it is the best theory *given all possible evidence*. As Smart puts it, the Ideal Theory is one which satisfies 'all possible operational and theoretical constraints', where the former comprise 'infinitely many observations carried out all over the physically accessible universe' (see also [6] pp. 472–3). Melchert's conception of what makes a story 'ideal' may correctly represent what certain anti-realists have in mind when they characterize truth as an idealization of warranted assertibility, but it is decidedly not what is at issue in the debate over metaphysical realism as that debate is understood by Putnam and Smart. (One might, of course, *dispute* this understanding by questioning whether the metaphysical realist *is* committed to the possible falsity of a theory which is 'Ideal' in Putnam's sense [3]. But Melchert is not canvassing *this* sort of objection.)

Let me briefly remark the difference between the preceding response to Melchert and a *prima facie* similar response by Peter Smith [12]. Smith's response rests crucially on the claim that

> Metaphysical realism as understood by Putnam and Smart turns on the doctrine that *an ideally well supported general theory of the world may still be quite radically wrong in its overall structure and ontological presumptions.*

We may talk of 'radical deception' where we have such erroneous belief concerning 'the basic general character of the world', contrasting this with cases of 'simple deception' where we have erroneous belief concerning particular features of the world, such as the proper interpretation of a given historical event. Smith's contention, then, is that one cannot argue from the possibility of *simple* deception in the face of the best available evidence to the possibility of *radical* deception in the face of such evidence. Given the above claim that the metaphysical realist needs to establish the *latter* possibility, and given that Melchert's example only establishes the *former* possibility, Smith concludes that 'Melchert's argument cannot settle the issue of metaphysical realism as standardly understood'.

I agree with Smith's conclusion, but I think it is misleading to argue for that conclusion by taking it as given that only cases of *radical* deception in the face of the best available evidence are relevant to establishing the correctness of metaphysical realism. It *is* true, I recognize, that the metaphysical realist is committed to the possibility of radical deception, and that most of the debate concerning metaphysical realism has focused on such a possibility. But these facts reflect more fundamental considerations – the metaphysical realist view of truth as 'transcendental correspondence', on the one hand, and the characterization of an 'Ideal Theory' as one which satisfies all *possible* evidence and not merely the best *available* evidence, on the other. To make his case, the metaphysical realist must provide an instance of a belief satisfying the following conditions: (i) we cannot deny the possibility that the belief is true (false), and (ii) we can account for the possible truth (falsity) of the belief only by appeal to the metaphysical realist model of truth as 'transcendental correspondence'. The problem with Melchert's example, I have suggested, is that it fails to satisfy the second of these conditions. It may be contingently true that *all* cases of *simple* deception fail to satisfy the second condition for similar reasons – all such cases may readily succumb to an anti-realist analysis in terms of assertibility in the face of *all possible evidence*. In this case, it will indeed be true that only cases of *radical* deception will suit the metaphysical realist's purposes. But we must still recognize that if we *could* come up with a case of simple deception satisfying both of the above conditions, this *would* establish the need for a metaphysical realist conception of truth, a conception which, in turn, would seem to entail the possibility of *radical* deception. In this mediated fashion, then, there *is* a route from the possibility of simple deception to that of radical deception.

We might try to resurrect Melchert's example by arguing that it *can* satisfy condition (ii) above – that even a story which was 'Ideal' in Putnam's sense would falsely maintain that the speaker's remark was intended as an insult. But this is not a promising strategy. For such an Ideal story would presumably be grounded in at least (a) the original speaker's introspective report of his intentions, (b) evidence, from the speaker's linguistic performance in contexts open to public assessment, of his understanding of the terms employed in making that report, and (c) an 'ideal' psychological theory comprising, *inter alia*, an account of the conditions under which a speaker may engage in self- or other-deception. It seems likely that such a story would 'truly' (*ex hypothesi*) assert that the remark was not insulting in intent. Suppose, though, that it asserted the contrary? It isn't clear what sense we could make, in this case, of our hypothesis that the remark was *not* intended as an insult, given the evidential grounding for the Ideal story. In fact, any such attempt to preserve the force of Melchert's example will be misguided if, as I shall suggest in response to Smart, all such examples proposed as counters to Putnam's argument against realism fail for a more fundamental reason. This will become apparent if we turn to Smith's attempt to answer Smart.

III

Smart asks us to consider the following case. We have two theories, T_I and T_I^*, which agree in their accounts of our four-dimensional space-time manifold, and which satisfy all possible operational constraints relative to that manifold. The theories differ only in that T_I^*

> asserts that our space-time manifold is a cross-section of a five-dimensional manifold which contains as another cross-section another space-time manifold, also, like ours, containing galaxies, stars, planets, and so on. Let us suppose that according to T_I^* we can never know about this other space-time sub-universe because this sub-universe and our own are precluded by the laws of nature from causally interacting with one another.

Smart argues that, given the theoretical constraint of simplicity, it would never be rational to prefer T_I^* to T_I, but that, nonetheless, T_I^* might be true, since the universe might be less simple than we anticipate.

Smith questions whether this line of reasoning can establish the possible *falsity* of an Ideal Theory. What facts does T_I get *wrong* if T_I^* is true? The obvious answer is that T_I mischaracterizes the dimensionality of the physical universe. But, Smith argues, we cannot give any content to the claim, in T_I^*, that the separate four-dimensional sub-universes are 'genuinely embedded in a physically real five-dimensional manifold' if, as T_I^* also maintains, the sub-universes are causally isolated from one another. And if we allow for causal interaction between the sub-universes, the 'empirical ideality' of T_I becomes moot. If, on the other hand, we pare down Smart's example to the 'key idea' that there may be causally unrelated sub-universes (or, more generally, 'empirically inaccessible regions of the world'), then, insofar as T_I observes the plausible theoretical constraint that a theory eschew 'surplus content', our Ideal Theory T_I will refrain from making *any* claims about the existence or nature of other sub-universes. In this case, the sort of possibility envisaged by Smart will not demonstrate that T_I might be *false*, but only that it might be *incomplete*.

Can this strategy rescue Putnam from Smart's objection? *Prima facie*, it seems that it can. Suppose we term the claim that belonging to an Ideal Theory is *sufficient* for the truth of a statement the 'weak' thesis, contrasting this with the 'strong' thesis that belonging to an Ideal Theory is both sufficient *and necessary* for the truth of a statement. Then Smith's claim is that Smart's example may count against the strong thesis but not against the weak thesis. Since Putnam's argument against metaphysical realism rests upon (a) the premiss that the latter requires the denial of the weak thesis, and (b) the 'model-theoretic' argument *for* the weak thesis, Smith's claim, if correct, would seem to completely defuse Smart's example as an objection to Putnam. But can Putnam afford to grant the force of Smart's example against the *strong* thesis? If we recall the details of the 'model-theoretic' argument, I think it will be apparent that he cannot do so.

Metaphysical Realism, for Putnam, is a purported model of the relationship between any correct theory and all or part of THE WORLD, where the latter comprises some fixed totality of mind-independent entities. The metaphysical realist assumes that there is a determinate and unique relation of reference between terms in a language and elements, or sets of elements, in THE WORLD. Truth, then, is taken to consist in a correspondence relation, mediated by reference so construed, between sentences in a language and states of THE WORLD. Putnam maintains [7] that metaphysical realists have assumed that a distinguishing feature of their model of truth is that truth is 'radically non-epistemic', in the sense that even an Ideal Theory might in fact be false. Putnam argues that the realist is *not* entitled to this assumption. His argument, formulated from *within* the realist's own framework, runs roughly as follows. Assume a metaphysical realist conception of THE WORLD and of truth as reference-mediated correspondence to THE WORLD. Then, for any Ideal Theory of an appropriate cardinality:

(a) There is a mapping, and hence a 'reference' relation, from the terms of the language of the theory onto elements, or sets of elements, in THE WORLD, such that THE WORLD satisfies the theory – call this mapping SAT; and
(b) Nothing in our *use* of the language can serve to exclude this mapping as an 'unintended' interpretation, since the Ideal Theory satisfies, *ex hypothesi*, all possible operational and theoretical constraints, and since SAT also satisfies our general intention to refer in such a way that our beliefs come out *true*.

Thus, if truth consists in a reference-mediated correspondence to THE WORLD, an Ideal Theory cannot fail to be true.

Crucial to Putnam's argument is the contention that there is nothing in the metaphysical realist's model that is able to ground the sort of unique, determinate reference relation necessary if we are coherently to entertain the possibility that an Ideal Theory might be false. Neither our general use of the language in which the theory is formulated, nor our espousal of a 'causal' theory of reference, nor (as Putnam elsewhere [9] argues against the likes of Michael Devitt and Davis Lewis [2] [4]) causal structure or reference relations 'in THE WORLD', can ground such a reference relation, according to Putnam. It is for this reason that our intention that our beliefs be true can validate all and only SAT-type interpretations as truly 'intended'; and, as Putnam establishes by the very construction of the SAT relations, on no such 'intended' interpretation can the Ideal Theory turn out to be false. (Note that it is for this reason that Putnam's argument does not equally well prove that an Ideal Theory must be *false* by some NON-SAT interpretation; thus Gregory Currie's claim [1] that Putnam's argument would establish both the truth and the falsity of theoretical statements and is therefore properly seen as an argument for scientific instrumentalism, is incorrect.)

What, then, if Putnam were to grant the force of Smart's example against the strong thesis? To grant this is to allow that certain statements might be true though not contained in an Ideal Theory. The obvious question, now, is, in virtue of *what* are such statements true? The truth of statements which *are* contained in the Ideal Theory is given, on Putnam's account, by the SAT relations which are the 'intended' interpretations of the langauge. Let S represent a statement such that: (i) S is true, and (ii) neither S nor not-S is contained in the Ideal Theory T_I. There is, indeed, an interpretation of $T_I \cup S$, call it SAT*, on which both T_I and S come out true. But there is also an interpretation of $T_I \cup$ not-S, call it SAT**, on which T_I comes out true and S comes out false. But since, *ex hypothesi*, we remain agnostic about S in our Ideal Theory, our intention that our beliefs come out true cannot select SAT* over SAT** as the 'intended' interpretation of our language. Thus the truth of S cannot be explained in the same way as the truth of those statements contained in the Ideal Theory. It seems, in fact, that we must say that if S has a determinate truth-value, it is in virtue of its corresponding to 'the way THE WORLD is', independently of whatever SAT relations may obtain. And *this* requires that there be a unique and determinate relation of reference between at least *some* of the terms in the language – those employed in S – and entities, or sets of entities, in THE WORLD. But, if such a relation exists and is a ground for the truth of *some* statements expressible in the language, we must surely draw the further conclusion that it is *this* relation, and not SAT, that is the genuine 'intended' interpretation of the language as a whole. And then there is no reason why statements contained in the Ideal Theory itself should not be false on this interpretation, even if true on SAT. Thus Putnam cannot grant the force of Smart's objection against the strong thesis. For, if he allows that an Ideal Theory may be *incomplete* with respect to certain expressible statements, then he must also allow that it may be *false*.

IV

I have argued that Putnam cannot avail himself of Smith's proposed strategy to defend himself against Smart. But this should give only momentary comfort to the cause of metaphysical realism; for, as may already be apparent, the very considerations which prevent Putnam from ceding the strong thesis to Smart also provide him with an answer to Smart's criticism. Smart believes that he can defend metaphysical realism against Putnam's 'model-theoretic' argument by producing a counter-example to the argument's *conclusion*, rather than by directly addressing the argument itself. But the counter-example can be made out only if Putnam's argument is flawed. Simply to offer the 'counter-example' as a refutation of Putnam is thus to beg the question at issue.

For consider what is involved in Smart's claim that T_I^* might be true. He maintains that the conceptual resources of mathematics and physics will permit us to describe the five-dimensional manifold postulated by the theory (or, perhaps, the alternative sub-universes which the theory envisages). But the truth-values of the sentences in T_I^* which incorporate these descriptions depend, according to the metaphysical realist, upon referential relationships obtaining between the relevant terms and entities, or sets of entities, in THE WORLD. Such sentences can have a determinate truth-value only if the relevant terms uniquely and determinately refer. But Putnam's 'model-theoretic' argument purports to show that there is nothing in the metaphysical realist model to fix the referents of the terms in a language, save the considerations adduced in the formulation of SAT, and this will apply to the terms in T_I^* as much as to any other terms. If Putnam's *argument* is correct, then the cosmological claims in T_I^* can be supposed to have a determinate truth-value only by way of an 'intended' interpretation SAT. But SAT presupposes the intention that one's beliefs be true, and this, in turn, requires that one *has* the beliefs in question. And, *ex hypothesi*, our Ideal community of inquirers does *not* believe the cosmological speculations of T_I^* – in fact, as Smart points out, the rational belief is that these speculations are *false*. So, if statements can have a determinate truth-value only by way of the interpretation SAT, the speculations of T_I^*, if they have a determinate truth-value at all, will be *false*. Such speculations can be supposed to be true only if Putnam's strictures against the possibility of SAT-independent determinate reference within the metaphysical realist framework are rejected. There are those, cited above, who have argued that these strictures *should* be rejected. But, until such arguments are made out, Smart's 'counter-example', and, indeed, any other example of this sort addressed to the *conclusion* of Putnam's argument, is question-begging. And, one might add, if such arguments *are* made out, Smart's example and others like it will be unnecessary; for, if there *are* independent constraints on reference of the sort urged by Devitt [2] and Lewis [4], then *any* claim in an Ideal Theory that transcends what is strictly dictated by operational constraints may be false.

University of Western Ontario,
London, Ontario, Canada N6A 3K7

© DAVID DAVIES 1987

REFERENCES

[1] G. Currie, 'A Note on Realism', *Philosophy of Science*, 49.2 (1982), 263-7.
[2] M. Devitt, 'Realism and the Renegade Putnam', *Nous*, 17.2 (1983), 291-301.
[3] J. Koethe, 'Putnam's Argument Against Realism', *Philosophical Review*, 464 (1979), 92-9.
[4] D. Lewis, 'Putnam's Paradox', *Australasian Journal of Philosophy*, 62.3 (1984), 221-36.
[5] N. Melchert, 'Metaphysical Realism and History', *Analysis*, 46.1 (January 1986), 36-8.
[6] H. Putnam, 'Models and Reality', *Journal of Symbolic Logic*, 45.3 (1980), 464-82.
[7] H. Putnam, 'Realism and Reason', in *Meaning and the Moral Sciences*, Boston: Routledge and Kegan Paul, 1978.
[8] H. Putnam, *Reason, Truth, and History*, Cambridge: Cambridge University Press, 1981.
[9] H. Putnam, 'Why There Isn't a Ready-Made World', *Synthese*, 51 (1982), 141-67.
[10] J. J. C. Smart, 'Metaphysical Realism', *Analysis*, 42.1 (January 1982), 1-3.
[11] P. Smith, 'Smart's Argument for Realism', *Analysis*, 43.2 (March 1983), 74-8.
[12] P. Smith, 'Metaphysical Realism and Historical Interpretation', *Analysis*, 46.3 (June 1986), 157-8.

RIGHTS, CONSTRAINTS AND TRUMPS

By PHILIP PETTIT

IN *Anarchy, State, and Utopia*, published in 1974, Robert Nozick argued that rights should be understood as constraints [2]. In *Taking Rights Seriously*, published three years later, Ronald Dworkin maintained that we should conceive of rights as trumps [1]. How do these claims relate to one another? Are they incompatible? Are they consistent but distinct? Or are they perhaps equivalent?

I wish to argue that they are equivalent. Although the claims are associated with rather different normative theories of rights, they amount to one and the same analytical thesis.

My paper is in three sections. In the first I show that if a right is a constraint then it is also a trump and that if it is a trump it is also a constraint; the two claims are equivalent. In the second I distinguish two propositions associated with each claim, suggesting that the constraint formula highlights one, the trump formula the other. And then in the final section I identify what I believe is the motivating ground for the two propositions: the belief that rights serve their bearers in an essentially protective role.

I CONSTRAINTS ARE TRUMPS, TRUMPS CONSTRAINTS

Robert Nozick introduces his conception of rights by setting up a contrast between rights considered as goals and rights considered as constraints. If the right to *X* is considered as goal, then one is obliged to violate *A*'s right to *X* when doing so increases non-violation overall: that is, non-violation by people generally of the right to *X* on the part of people generally. If it is considered as a constraint, then this is not so. On the contrary, one is obliged not to violate *A*'s right even in such a circumstance. (It is significant that the non-violation which one is not allowed to promote under this conception of rights – and under Dworkin's – is non-violation by people generally (not by oneself) and of people generally (not of the holder of the right). That one may violate *A* in order to promote non-violation by oneself is intelligible because in such a case the rights with which one is faced are inconsistent. That one may violate *A* in order to promote non-violation of *A* is intelligible if, as is argued in section 3, the point of rights is to protect the interests of those who hold them.)

The constraint conception of rights means that one sees *A*'s right to *X*, not as part of an overall goal – the reduction of violation by anyone of anyone's right to *X* – but as something which binds one regardless of how that goal is served. This conception does not entail, notice, that rights bind the agent regardless of how all other goals are served; strictly, it says nothing on the status of rights vis-à-vis independent aims. But as it happens Nozick endorses the claim that rights have priority over all goals, at least under normal circumstances, and he often presents the constraint conception of rights in the context of that claim. 'The side-constraint view forbids you to violate these moral constraints in the pursuit of your goals; whereas the view whose objective is to minimize the violation of these rights allows you to violate the rights (the constraints) in order to lessen their total violation in the society' ([2] p. 29).

Ronald Dworkin introduces the conception of rights as trumps by focusing on the role which rights play – or, as he thinks, ought to play – in controlling the demands of collective goals. A collective goal, typically the sort of thing that governments pursue, is an aggregate social benefit: economic efficiency, equality in opportunity, or whatever. An individual's rights, on Dworkin's conception, serve to protect her from certain sorts of treatment which the untrammelled promotion of such a goal might require. 'Rights act as trumps over some background justification that appeals to the collective welfare' ([1] p. 367).

I maintain that constraints are trumps and, conversely, that trumps are constraints. In defending this thesis, and in bolstering it in the next two sections, I will be offering further explication and elaboration of Nozick's and Dworkin's views. I realize that the characterization just given is inadequate in a number of ways.

First, then, to the claim that constraints are trumps. A constraint such as that which the right to *X* is meant to provide serves to protect the individual from the demands of the following collective goal: the reduction of violation by people generally of the right to *X* on the part of people generally. In this respect therefore such a constraint is a trump held by the individual which he can use to control the demands of that goal.

Nozick thinks, as I have suggested already, that the constraint provided by a right will do even better than this by the individual. He thinks that it will serve to trump the demands of any goal and therefore the demands of any collective goal. A constraint in this view of things constitutes a total trump, as we might say. Or at least a total trump under normal conditions, for Nozick suggests that rights may be suspended when there is a threat of moral catastrophe ([2] p. 30n).

So far as a right is a constraint, then, it must also be a trump, whether a partial or total trump. But what of the converse claim? Does the fact that a right is a trump mean that it is also a constraint?

If the right to *X* is presented as a trump, then that means that it protects the individual from the demands of at least certain collective goals. As Dworkin says, 'no alleged right is a right (on my account) unless it overrides at least a marginal case of a general collective justification' [11] p. 366). The question then is whether such a trump necessarily overrides the demands of the following collective goal: the reduction of the violation by people generally of the right to *X* on the part of people generally. If it does, then trumps are always constraints; if it does not, trumps may fail to be constraints.

On Dworkin's conception of trumps, they do always serve a constraining role. There are two pieces of evidence for this. One is internal to the characterization of trumps, the other is external to it.

The internal evidence is that trumps are meant to protect individuals, for they could scarcely be hailed for this feature if the goals they trumped did not include the goal of reducing overall violation of the right in question. If the right to *X* trumps certain other goals but is not a constraint in Nozick's sense, then it has the status of a collective goal which is given infinite weight against those other goals; otherwise put, it is a lexicographically prior collective goal. With such a status, the right to *X* may be held to protect *X*, whether *X* be liberty of speech, equality of opportunity, or whatever. It cannot so obviously be said to protect any particular individual however, since the promotion of the right to *X* may now require the denial of that right to this or that person.

The external evidence for saying that trumps in Dworkin's sense are constraints is clinching. It comes from a discussion of his conception of rights in which Dworkin varies his usual terminology. What he says is that *A* has the right to *X* so far as the goal of *A*'s enjoying *X* has a privileged status vis-à-vis certain other goals. The important thing is that the goal of *A*'s enjoying *X* has this status, not the goal of a maximum number of people's doing so. For if the goal of *A*'s enjoying *X* serves as a trump – we know independently that this is what the privileged status involves – then this means that *A*'s right to *X* is a constraint. 'A political right is an individuated political aim. An individual has a right to some opportunity or resource or liberty if it counts in favour of a political decision that the decision is likely to advance or protect the state of affairs in which he enjoys the right, even when no other political aim is served and some political aim is disserved thereby, and counts against that decision that it will retard or endanger that state of affairs, even when some other political aim is thereby served' ([11] p. 91).

I conclude that constraints are trumps and trumps constraints. There are important differences in the normative theories of rights which Nozick and Dworkin defend, since they diverge on the source of rights, the range of rights, and the related issue of whether or not rights should be seen as total trumps. But these differences do not affect their agreement on matters of definition. They defend an analytical conception of rights which is capable of being invoked under a variety of normative approaches (see [3] and [4]).

II THE TWO SIDES TO THE CONSTRAINT-TRUMP CONCEPTION OF RIGHTS

Our argument for the equivalence of Nozick's and Dworkin's conceptions of a right has brought towards the surface something which it is worth highlighting. This is that there are really two propositions associated with the thesis that rights are constraints or trumps. One of these propositions is given prominence by the metaphor of the constraint, the other by the metaphor of the trump.

The first proposition is that a claim such as the right to *X* is untradable across persons. It constitutes a distinct demand on the part of each bearer; it is, as we might say, a personalized claim. This proposition is explicit in the description of rights as constraints but it is also implicit, as we have seen, in their characterization as trumps. It means that the fact that violating *A*'s right to *X* would reduce violation by others of others' rights to *X* does not legitimize the act of violation. *A*'s right to *X* is not subject to such negotiation; it is incommensurable with other people's right to *X*.

The second proposition involved in the constraint-trump conception is that any right, or at least any right worth its salt, is not just untradable across persons, but also untradable against certain other goods. The right is given an infinite weight in competition with those goods; only when it is satisfied is one entitled to concern oneself with the promotion of such benefits. This proposition depicts rights as claims that are privileged as well as personalized. It is not strictly implied in the characterization of a right as a constraint or trump, since a claim will count as a constraint or trump just in virtue of being personalized. But the proposition is strongly enough associated with the conception – and in particular with the trump formulation – to be identified as a component of it.

The conception of rights shared by Nozick and Dworkin may be summed up then as the view that rights are simultaneously personalized and privileged claims. These two elements are clearly present in the assumptions of each, as our discussion in the last section will have brought out. And yet the elements are equally clearly distinct. A claim might be personalized without being privileged in any way against the requirements of other goods. And a claim might be privileged to one or another extent without being personalized: it would amount in that case to an aggregate good which was given infinite weight or lexicographic priority in competition with the goods over which it was privileged.

One advantage of explicating the shared conception of rights in this way is that it also enables us to pinpoint the normative divergence between Nozick and Dworkin. Both agree that all rights must be privileged in some degree but they take different views of the extent of the privilege that is actually involved. For Dworkin it is such that the right trumps a certain unspecified number of collective goals; for Nozick it is such that the right trumps all.

III THE UNITY OF THE CONSTRAINT-TRUMP CONCEPTION OF RIGHTS

The fact that their conception of rights has these two distinct elements raises the question whether Nozick and Dworkin are not running two essentially distinct conceptions together. I would like to conclude by arguing that they are not. I believe that the two elements fall out of a common and unifying notion of the role that rights do or ought to play in relation to their bearers.

A person's rights are on a par with her desires, or needs, or merits, so far as they constitute considerations that others are expected to take into account in certain sorts of behaviour towards that person. Like those other factors her rights constitute claims whose satisfaction is in the interest of the bearer in some sense.

But claims like rights, desires, needs and the like may serve a person's interests in one of at least two different ways. They may be designed to ensure that others take account of the bearer's interests in their deliberations or, more strongly, in their doings: not all their doings perhaps, but at least doings that occur under certain conditions. I may consider someone's interest in my deliberations and yet, for reasons of overall optimization, override them in what I ultimately decide to do. If I am required to take account of those interests not just in how I deliberate, but in what I actually do – universally or under certain conditions – then I have no such leeway. I must cater for the person's interests in the minimal measure required if, intuitively, they are to be satisfied.

Claims which ensure that a bearer's interests are taken into account at the level of deliberation, and only at that level, serve, we may say, to promote the person. Claims which ensure the more substantial response serve, in contrast, actually to protect the bearer: specifically, to protect her from all or some actions on the part of others that would be inimical to the interests involved. Claims of the promotional sort guarantee a certain sort of consideration for the bearer; claims of the protective sort guarantee a certain sort of treatment.

I believe that what unifies the constraint-trump conception of rights is the assumption that rights serve their bearers in a protective rather than just a promotional role. The reason is that if a person is to be protected rather than just promoted, then it is natural that she be credited with claims that are at once personalized and privileged. This can be fairly readily demonstrated.

If *A*'s claim to *X* serves her in a protective role then that means that there is a reason for others, in the absence of the claim, not to provide her with *X*. The reason against which the claim protects *A* will be *X*-related or *X*-unrelated. It will arise from the fact that providing *X* for *A* militates either against the goal of reducing overall violation of people's claim to *X* or against some independent goal, *G*. Through being personalized, *A*'s claim to *X* will protect her against the first sort of reason; through being privileged, it will protect her against the second, or at least against certain forms of the second.

This does not quite show that it is natural to see rights as personalized and privileged if one thinks that they have a protective role. We also need to establish that protection cannot be provided by claims of a weaker sort.

The need is pressing. for it may seem that *A* can be protected against the influence of both *X*-related and *X*-unrelated reasons by something less than the fully personalized and privileged claim. Suppose she had a claim which protected her against the promotion by however small a margin of the *X*-related and *X*-unrelated goals but which did not protect her against the possibility of promoting either goal by some substantial degree, *K*. Wouldn't such a part-personalized and part-privileged claim still provide protection as distinct from promotion?

It would. But that fact is no objection, since every part-personalized claim is equivalent to a fully personalized counterpart, every part-privileged claim equivalent to a fully privileged counterpart.

The part-personalized claim to *X* is equivalent to the claim to benign *X* – *BX*, for short – where *X* ceases to be benign when its enjoyment means that the goal of reducing overall violation of the claim to X is sacrificed to degree K or beyond. And if the claim to X is part-personalized, then the claim to BX is fully personalized. For consider the situation if a person can invoke a claim to BX – she can enjoy X without causing a disadvantage of degree K – but the claim is not fully personalized. In such a situation the person may be denied X for the sake of a benefit to others which is less than degree K. And that means that in that situation not only is the claim to BX not fully personalized; neither is the claim to X part-personalized. Thus, if the claim to X is part-personalized then the equivalent claim to BX must be fully personalized.

A similar argument shows that a claim to X which is part-privileged with respect to a goal, G, is equivalent to a fully privileged counterpart. The fully privileged claim will be to an X that is benign from the point of view of G – GBX, for short – where X loses such benignity when it means a sacrifice of G to degree K or more.

I conclude that given the view that rights play a protective role, it is quite natural to see them as at once personalized and privileged claims. Despite its distinct aspects then, the constraint-trump conception of rights has the sort of unity that we ought to require of a satisfactory analysis.[1]

Research School of Social Sciences,
Australian National University,
Canberra, A.C.T. 2601, Australia

© PHILIP PETTIT 1987

REFERENCES

[1] R. Dworkin, *Taking Rights Seriously*, new impression, London: Duckworth, 1977, pp. 90-4 and 364-8.
[2] R. Nozick, *Anarchy, State, and Utopia*, New York: Basic Books, 1974.
[3] P. Pettit, 'Can the Welfare State Take Rights Seriously?', in D. Galligan and C. Sampford, eds, *Law, Rights and the Welfare State*, London: Croom Helm, forthcoming.
[4] P. Pettit, 'The Consequentialist Can Recognise Rights', *Philosophical Quarterly*, forthcoming.

---

[1] I benefited from comments received from Alan Hamlin and from an anonymous referee and am grateful to both. I wrote the final version of the piece when I held an Overseas Fellowship at Churchill College, Cambridge and I am also grateful to the College for the facilities with which I was provided.

DESERT AND ENTITLEMENT: A RAWLSIAN CONSEQUENTIALIST ACCOUNT

By DAVID CUMMISKEY

I RAWLS, NOZICK, AND FEINBERG ON DESERT AND ENTITLEMENT

RAWLS claims that we have no natural entitlement to the benefits generated by superior natural talents and abilities. To allow distributive shares of primary social goods to be determined by natural talents and abilities, he argues, is to let people's life prospects be determined by factors that are arbitrary from a moral point of view. We should, instead, agree 'to regard the distribution of natural talents as a common asset and to share in this distribution whatever it turns out to be' ([3] p. 101). I will refer to this as 'the common assets thesis'. Nozick responds by arguing that 'it needn't be that the foundations underlying desert are themselves deserved, *all the way down*'. He claims that we just may *have, not illegitimately*, some of the things or characteristics we use in coming to deserve something ([2] p. 225). Our natural talents and abilities are simply things we have, not illegitimately. And, since it is we that have them, why, he asks, are we not entitled to the benefits that are generated by them? It does not follow from the moral arbitrariness of this distribution that it may be treated as a common asset. Indeed, the very considerations supposed to show that we are not individually entitled to these assets would also show that we are not jointly entitled to the benefits. The natural distribution may be arbitrary and undeserved, but it is nonetheless the natural distribution. If no one has a claim on the pool of natural assets, then why should not those who fortuitously have these talents and abilities benefit from them?

Nozick's thesis, about the foundations of desert, is rather plausible and it does provide *some* support for his conclusion that those who fortuitously have the talents and abilities are entitled to the benefits that flow from them. Nonetheless I will argue that Rawls can, and does, accept the thesis that the foundations of desert need not be themselves deserved, and that Rawls can meet whatever burden of proof this shifts on to him. In order to demonstrate both the accuracy of Nozick's thesis about the foundations of desert and my claim that it is acceptable to Rawls, let us look more closely at the concept of desert.

Joel Feinberg has argued (i) that all attributions of personal desert must have a basis, and (ii) that the basis of a subject's desert must always be some factor about that subject. These two theses help clarify Nozick's thesis. (i) First, 'it is necessary that a person's desert have a basis' ([1] p. 61). 'Desert without a basis simply is not desert' ([1] p. 58). This feature of desert is obvious if one reflects that it makes no sense to claim that someone deserves something 'for no particular reason'. As Feinberg puts it, 'judgments of desert carry with them a commitment to the giving of reasons' ([1] p. 58). The reasons that must accompany judgements of desert are the basis of the alleged desert. (ii) Second, it is a necessary condition that 'the facts which constitute the basis of a subject's desert must be facts about that subject' ([1] pp. 59, 61). 'If a person is deserving of some sort of treatment, he must, necessarily, be so *in virtue of* some possessed characteristics or prior activity' ([1] p. 58). Shawn cannot deserve a good grade simply because Cedric did well on the test. If Cedric's performance is the basis of Shawn's desert, then Cedric's performance must somehow be a reflection of Shawn's abilities or prior activities.

With this (partial) analysis in hand we can see what is plausible about Nozick's thesis. Nozick is simply pointing out that the possessed characteristics or prior activities, which provide the basis of someone's desert, need not themselves be deserved. If I win a race because of my natural talent, training, and tactical ability, then I may plausibly claim to deserve the prize. My personal desert does not presuppose that I also deserved my talent, my ability to train, and my tactical ability. As the concept of desert is ordinarily understood, it simply is not the case that one must deserve the basis of one's desert. Indeed, if the bases of desert, the talents and abilities, had to be deserved, then they too would have to have a basis which too would have to be deserved; etc. It would seem that the resulting regress would make desert impossible.

Nonetheless, although I agree with Nozick's thesis about the foundation of desert, I fail to see how it helps his case. Nozick's thesis is consistent with both an institutional and a pre-institutional notion of desert. Rawls grants that within the contexts of legitimate social institutions, those who do what the institutions announce they will reward deserve their rewards ([3] pp. 10, 103). And Rawls believes that public institutions which involve social and economic inequalities often provide incentives which work to the advantage of the least advantaged, and which are thus justified. Since he does not think that the bases of these entitlements are themselves deserved, he accepts Nozick's thesis; it is a corollary of his own position.

Nozick does, nonetheless, have a point. I would grant that the above objection shifts the burden of argument back to Rawls. Natural talents and abilities are not 'manna-from-heaven': even if the distribution of natural talents and abilities is arbitrary, these natural 'assets' are nonetheless distributed.

> Since things come into being already held (or with agreements already made about how they are to be held), there is no need to search for some pattern for unheld holdings to fit ... The situation is not an appropriate one for wondering, 'After all, what is to become of these things, what are we to do with them?' ([2] p. 219).

Since holdings come into the world already held, is it not plausible to assume that those who fortuitously have the holdings are pre-institutionally entitled to the benefits that flow from them? We can legitimately treat natural talents as a common asset only if there are no pre-institutional entitlements to the benefits that flow from them.

II THE LOGIC OF DESERT AND ENTITLEMENT: A RAWLSIAN CONSEQUANTIALIST RESPONSE

By looking at the difference between desert and entitlement and their essential connection to social institutions, we will see that pre-institutional accounts of these notions are inadequate. *Desert bases are determined by antecedent goals,* and thus Nozick's entitlement theory is essentially flawed. Since Nozick does not present and defend a pre-institutional account of these notions I focus on Feinberg's argument.

The difference between desert and entitlement has led Feinberg mistakenly to conclude that 'desert is a *moral* concept in the sense that it is logically prior to and independent of public institutions and their rules' ([1] p. 87). Although Feinberg's analysis of the difference between the two concepts is correct, his conclusion does not follow.

Feinberg begins by pointing out that 'to say a person deserves something is to say that there is a certain sort of propriety in his having it' ([1] p. 56). He then attempts to clarify 'the particular kind of propriety distinctive of desert ... by contrasting it with other forms of propriety' ([1] p. 57). He distinguishes three kinds of propriety conditions. 'There are those whose satisfaction confers eligibility ("eligibility conditions"), those whose satisfaction confers entitlement ("qualification conditions"), and those conditions whose satisfaction confers worthiness or desert ("desert bases")' ([1] p. 58). In the case of the office of the presidency of the United States the eligibility conditions are natural-born citizenship and a minimal age of thirty-five years. The entitlement condition is getting the most electoral votes in a fair election. If one satisfies both the eligibility conditions and the entitlement conditions then one 'can claim the office ... as his *right*, according to the rules he is entitled to it' ([1] p. 57). Satisfaction of these conditions does not, however, confer worthiness or desert.

> To deserve something, one must be qualified in still a third sense: one must satisfy certain conditions of worthiness which are written down in no legal or official regulation. Thus to be 'truly qualified' for the presidency, a person must be intelligent, honest, and fair minded; he must have a program which is really good for the country and the tact and guile to make it effective ([1] p. 57-8).

We thus have a distinction between entitlement or victory conditions and worthiness or desert bases. Even though ordinary language does not precisely distinguish these two concepts, the difference is familiar and intuitive. In many cases of awarding prizes, grades, rewards, offices we have entitlement without desert or desert without entitlement. The person who satisfies the entitlement conditions need not be the same as the one who is most deserving.

Granting all of this, in what sense does it follow that 'desert is a *moral* concept . . . that is logically prior to and independent of public institutions and their rules'? To answer this question we must diagnose what happens when desert and entitlement come apart.

In the simple case of a contest of skill for a prize it often occurs that the person with the most skill fails to satisfy the victory condition. When this occurs, the most likely cause is some stroke of bad luck which leads to an off day. The next likely cause is that the victory conditions themselves were in some way ill chosen. Given this diagnosis of the likely causes, does it follow that desert is 'logically prior to and independent of' the race and its rules?

A race is a social institution which is usually constructed for some specific purpose. We cannot conclude that someone deserves to win the race without, at least tacitly, appealing to the purpose of the race. When we say that the most skilled deserves to win then we are assuming that the point of the race is to reward skill. Since rewarding skill is the point of the institution it also is the appropriate basis of desert. In an important sense desert is not independent of the rules, for the point of the rules is to single out the one most deserving of victory. If the winners of a race were habitually not the most deserving *because* the rules were inadequate, then, if possible, the rules would be changed so as to bring entitlement and desert more closely together. In this sense desert is not independent of public institutions and their rules.

Desert is 'logically prior to' public institutions and their rules only in the sense that desert is directly related to the point of the institutions and their rules. Indeed, desert bases are determined by the point of the institutions. Without the prior social ends the institutions would have no point and there would thus be no appropriate basis of desert. The logical relations between a social goal or end, desert, and entitlement are as follows. First, we have the social goals or ends. These may be moral, immoral, or amoral. Indeed, this analysis is compatible with absolutism, relativism, or non-cognitivism. Next, we have desert bases. The appropriateness of desert bases is determined by whether or not a social policy of praising or blaming a subject with some characteristic or because of some prior activity is likely to further the goal of the institution in question. Last, we have entitlement conditions. These are the social mechanisms and social practices which reward or sanction individuals with the appropriate desert basis. In essence, entitlements set up expectations of specific treatment because of specific prior activities or specific personal characteristics. Desert is logically prior to institutions in the same way that the point of the institutions is prior to the institutions. Feinberg has not established that desert is logically prior in any sense that would provide a constraint on the legitimate construction of institutions.

In the above example of the race I have followed Feinberg, and others, and argued by analogy from a fairly simple case. The example suffices to show that Feinberg's conclusion does not follow from his argument. Nonetheless, to avoid my conclusion, one might argue that I have illegitimately assumed that the case of the race and cases of moral desert are analogous. I respond as follows. First, the example of the race provides a clear and non-mysterious analysis of desert. If one accepts the analogy, one has an equally clear and non-mysterious analysis of moral desert. Second, this analysis provides the basis for a plausible account of the nature of backward-looking reasons, which are constitutive of some institutions. It thereby lays the foundation for an account of backward-looking reasons within a larger consequentialist framework. I believe that the compatibility of the analysis with both consequentialism and Rawls's Kantian theory is a virtue. Third, if one accepts my analysis of desert and entitlement for the case of the race but argues that moral desert is disanalogous, then one is committed to a non-univocal analysis of desert. *Prima facie*, however, the only difference between moral desert and other desert claims is that moral desert involves the moral point of view, the giving of moral reasons. Fourth, both Rawls and Nozick, in the final analysis, wish to avoid intuitionist appeals. Thus, within the context of this paper, the above objection assumes that there is a plausible non-intuitionist pre-institutional analysis of moral desert and entitlement. Until we have such an analysis in hand, we have no motivated reason for believing that the analogy is inadequate. The burden of proof has been shifted back.

Nozick's objection to Rawls presupposes a notion of pre-institutional desert and entitlement. I have argued that desert is an institution-dependent concept, and thus Nozick's objection to Rawls cannot be sustained.[1]

Bates College, Lewiston,
Maine 04240, U.S.A.

© DAVID CUMMISKEY 1987

REFERENCES

[1] J. Feinberg, *Doing and Deserving*, Princeton: Princeton University Press, 1970.
[2] R. Nozick, *Anarchy, State, and Utopia*, New York: Basic Books of Harper Torchbooks, 1974.
[3] J. Rawls, *A Theory of Justice*, Cambridge: Belknap Press of the Harvard University Press, 1971.

[1] I would like to thank Sarah Conly, Steve Darwall, Allan Gibbard, Mike Lavin, John Nolt, Betsy Postow, Peter Railton, and the referees of *Analysis* for helpful discussion or comments.

THE GENERALITY CONSTRAINT AND CONSCIOUS THOUGHT

By G. R. GILLETT

IN this paper I want to argue that there is a conceptual relation between 'The Generality Constraint' and the idea of a conscious thinking subject (as I have sketched in 'Brain Bisection and Conscious Experience', *Mind*, April 1986).

In describing 'The Generality Constraint', Gareth Evans states,

> ... if a subject can be credited with the thought that *a* is *F*, then he must have the conceptual resources for entertaining the thought that *a* is *G*, for every property of being *G* of which he has a conception. This is the condition that I call 'The Generality Constraint' (G. Evans, *Varieties of Reference*, J. McDowell, ed., Oxford U.P. 1980, p. 104).

By this remark and the foregoing discussion he directs us to the notion that in thought there is a need for the items of which we think and which enter into our complex thoughts to be able also to figure in a number of different but conceptually related thoughts. He considers that this feature sets apart mere responders or information processors from true conceptual thinkers.

> Even readers not persuaded that any system of thought must conform to the Generality Constraint may be prepared to admit that the system of thought we possess – the system that underlies the use of language – does conform to it. (It is one of the fundamental differences between human thought and the information-processing that takes place in our brains that the Generality Constraint applies to the former but not to the latter. ...) (Evans, op. cit., p. 104, fn 22).

It is this last claim that I want to both defend and use to illumine an aspect of conscious thought.

It is clear that *if* we have a subject who can say of, say, an envelope that it is yellow but when asked whether there is anything else in the room that is yellow cannot identify as being such a sweatshirt, a timetable, or a book, *then* we could not say of him and he could not say of himself that he had grasped the concept ⟨yellow⟩. When he could correctly pick out the feature in a number of its instances or apply the concept to a number of objects to which it can be correctly applied, then we and he would know what he meant by 'yellow' – he would have a grasp of the concept ⟨yellow⟩. Note that this is not an argument based on our need to verify that he has the concept, i.e. on what evidence would serve to warrant an assertion that he possessed the concept, but is rather a fact about what is constitutive of having the concept in question. It highlights for us the basis of possession of the concept which is mastery of a certain practice of subsuming under judgements (I. Kant, *Critique of Pure Reason*, tr. N. Kemp Smith, Macmillan, 1928, B171). Wittgenstein would call this the 'grammar' involved in the meaning of 'yellow' (L. Wittgenstein, *Philosophical Investigations*, I.371, tr. G. E. M. Anscombe, Blackwell, 1953). We might convince ourselves of this by asking of the defective thinker who says 'the envelope is yellow' but then cannot go on to make any other judgements about yellow objects, 'What can he be thinking?'. Our justification for ascribing to him (or his for ascribing to himself) the conscious thought ⟨that envelope is yellow⟩ is deficient in respect of part of the content of that thought, namely the part that depends upon his ability to exercise related judgements in a series of thoughts which share the structural feature we indicate by the word 'yellow'.

We now need to ask whether such manifestability is in all respects an *a priori* necessity for the possession of the concept. (Here we need to be quite careful not to allow consideration of a particular situation where a concept user is impaired in some way to influence our decision about the general conditions for the existence of the object of our attention, in this case a conceptual ability.)

Imagine a thinker who notes a yellow item in his world. He may apply the concept ⟨yellow⟩ to the item or merely respond to that item in some limited way. It will only be as he reacts in such a way as to forge a connection with the colour of yellow items in a wide variety of situations that he will be exercising the conceptual ability associated with the word 'yellow'. If the mental act that he exhibits is such as to be able to be used in this general way with respect to a variety of possible instances or non-instances, he will be able to achieve and refine a grasp of the concept ⟨yellow⟩. Note what is involved here. Of all the aspects and possible responses available in the situation, only that act which can potentially be generalized in the requisite way will count as a conceptual act involving 'yellow'. To have those properties the putative conceptual act must effect a selection from the many possible aspects of a situation of just and only that aspect which counts as falling under the concept ⟨yellow⟩. This means that he must exhibit the kind of wilful directed attending which will equip him to deploy a range of abilities according to his changing interests and projects. In as much as the thinker is able to conceive of the same feature as common to a number of situations and the same object as the potential instance of a range of predicative judgements, so there must be this flexible 'aspect seeing' attentional control and selectivity of response. But this is just what it is to be conscious.

The point is reinforced by considering the acquisition of the ability which counts as the grasp of a concept. A natural responder such as a child has a tendency to react to its environment in certain very human ways. It also has the ability to exhibit a congruent response to other human beings. So it learns, for instance, to smile, as we ourselves can imitate another's expression or know when our own is not being imitated accurately without the aid of a mirror.

P4709

These natural reactions come to be constrained so as to conform to certain patterns of partly natural but essentially rule governed response which are exhibited by others and imparted to the child. Such responses are essentially (because they have a generative use) either inter-situational or intra-situational-feature-selective, and inter-subjective. When the child has latched on to an ability of this type we ascribe a grasp of the concept concerned and in so doing make available to the child the ability to ascribe to itself and think in a normative way about that concept. The normative determinants of the ability will incorporate the potential generality of its application and thus the structural features of a true concept. The entire process depends upon a congruence in attentional and behavioural abilities to allow a mastery of the human techniques governed by the rules which individuate conceptual elements (of course, such rules need not be discursively explicit). Evans notes,

> I should prefer to explain the sense in which thoughts are structured, not in terms of their being composed of several distinct elements, but in terms of their being a complex of the exercise of several distinct conceptual abilities (Evans, op. cit., p. 101).

In its totality this range of abilities to exercise informed aspect selection and generalization at will and then use the results of that faculty in an action guiding way is what constitutes consciousness. So Wittgenstein remarks,

> Consciousness is as clear in his face and behaviour as in myself (L. Wittgenstein, *Zettel*, G. E. M. Anscombe and G. H. von Wright, eds., Blackwell, 1967, p. 221).

In that this also and essentially makes conscious experience experience within a structure of thought, in congruence with the language-using community, it involves the subject of experience as an inhabitant of the 'space of reasons' (J. McDowell, in a graduate class on philosophy of mind, Oxford, Hilary Term, 1986).

Where does this leave a mere responder or information processor? Such an entity can only react so as its reaction within and across situations is keyed to a certain feature. The very fact that it lacks a flexible and structured ability to pick out this property and then that or this object instancing a given property and then that (actual or conjectural) object means that (a) it does not have a concept which is an element of a structured range of thoughts and (b) it is not conscious.

(a) It does not have the concept because it cannot exercise an ability to attend selectively to, or mentally isolate and identify, as an aspect of its response to a given situation, a feature which not only has multiple instances but also could figure as an element of one of a range of related thoughts all referring to a single subject of predication or object with diverse properties. A mere responder emits a response which is keyed to a triggering condition (or complex of such), in principle extensionally specifiable. By contrast, the situations where a thought may be exhibited need not share any extensional property; a thinker can just as well say 'It is not red' as 'It is red' in exercising her grasp of the concept ⟨red⟩, provided there is some human or intensional sense in the occasion of use. That ability or set of abilities in virtue of which the Generality Constraint is fulfilled by a subject of experience is just the attentional selectivity and imaginative creation of representations (by which linked terms or actions are used to direct the attention of others with congruent natural capacities) that constitutes a grasp of the concept ⟨red⟩. In this way a subject makes use of a large and intersecting repertoire of abilities in detecting and structuring her experience so that she can think about it.

> Light dawns gradually over the whole (L. Wittgenstein, *On Certainty*, tr. G. E. M. Anscombe and G. H. Von Wright, Blackwell).

(b) The stimulus-response information processor cannot direct its attention and its activity in the light of an interested and action-guiding appreciation of the divers possibilities of a situation. Its lack of the ability to focus and direct its attention and thus to exploit any of the features of a given situation in which it finds itself constitutes a lack of structural generality with respect to thought about a constant subject. The complementary lack of the ability to discern a constant feature across different situations which share some focus of directed interest or selected attention is just a lack of structural generality with respect to universals. If it were conscious, it would discern and exploit these possibilities in a flexible way according to patterns of interest and so exercise a structured set of abilities. The fact that it does not do this means that it is not conscious. All this creature does is to produce a given response when impinged upon by a canonical stimulus. The lack of refinement of this ability and interactive modification of it by the role it plays in a structured repertoire of abilities means that all that is required in its explication is an information processing network.

To summarize, the suggestion that there is a conceptual link between the structure of thought revealed by the 'Generality Constraint' and conscious experience has been pursued.

What is involved in the abilities which constitute thought structured according to the generality constraint is such that it requires a flexible selectivity of attention and response to situations. The mental acts involved allow thought with content rather than mere causal influences to figure in the ascriptions which can be made to a subject.

The ascription and self-ascription of thought with content results from that congruence of abilities and capacities between subjects which allows a structure of rule-governed acts to be evolved.

The joint facts (i) that there is a constitutive selectivity of attention and response, (ii) that there is congruence of activity which allows (in principle) conformity to rules shared by rational interlocutors, means that a thinker of structured thoughts is as clearly conscious and rational as those interlocutors. He is thus both equipped and entitled, in a way which is conceptually linked to his fulfilment of the 'Generality Constraint', to be a member of 'the kingdom of ends' (Kant) or 'the space of reasons' (McDowell).[1]

Magdalen College and Ian Ramsey Centre,
Oxford

© G. R. GILLETT 1987

---

[1] I would like to acknowledge the helpful comments and suggestions on this paper offered by Stephen Blamey and Kathy Morris.

ON PAN-CRITICAL IRRATIONALISM

By PAUL HELM

IN *The Retreat to Commitment* (2nd edition 1984; 1st edition 1964; all references are to the 2nd edition), William Warren Bartley defends what he calls 'pancritical rationalism' against the foundationalist epistemologies of classical rationalism and empiricism. The trouble with such foundationalism, according to Bartley, is that it involves a 'retreat to commitment' for it relies upon a person's non-reasoned commitment to certain foundational truths and thus has no answer to any irrationalist who may, in turn, make his irrationalism his basic commitment.

So the rationalist (and any foundationalist) and the irrationalist are in the same logical boat. Bartley calls this the *tu quoque* argument.

> (1) for certain logical reasons, rationality is so limited that *everyone* must make a dogmatic irrational commitment; (2) therefore the Christian has a right to make whatever commitment he pleases; and (3) therefore, no one has a right to criticize him (or anyone else) for making such a commitment (p. 72).

It is Bartley's aim to defeat the *tu quoque* argument 'by showing that it is possible to choose in a nonarbitrary way among competing, mutually exclusive theories' (p. 83).

Take classical rationalism, for example. According to Bartley it comprises two theses. The first is that any position which can be justified or established by rational argument is to be accepted, and the second is that *only* positions which can be justified or established by rational argument are to be accepted. 'The panrationalist accepts anything that can be rationally justified, and also is ready to justify rationally anything that he accepts' (p. 93). Bartley shows that it is impossible to hold these two requirements simultaneously, because the acceptance of the second requires that the first requirement be justified by rational argument, and this is impossible. And even if it could be achieved such an argument would only be convincing to someone who already accepted the need for justification, which is precisely the point at issue (pp. 93-4).

So which of the two ought to be rejected, since they are requirements that cannot be held compatibly?

> Now there are several good reasons for rejecting the second requirement rather than the first. Since we are searching for an adequate rationalist identity, we shall hardly want to abandon the demand that the rationalist accept any position that can be rationally justified. Moreover, the second requirement can be shown by argument to be self-contradictory ... because it, too, cannot be justified by appeal to the rational criteria or authorities (pp. 94-5).

So Bartley is convinced of the need for what he calls *critical* rationalism. It is not sufficient to acknowledge, as he thinks Ayer (p. 97) and Popper (p. 104) do, that some basic matters cannot, as a matter of logic, be justified, that what sets the standards cannot itself, as a matter of logic, be subject to those standards. Such an acknowledgement simply underlines the 'fideism' of traditional rationalism, a fideism which goes back, Bartley thinks, to the fact that rationalism offered itself as a rival authority to theology.

So Bartley makes a radical proposal to outflank the *tu quoque* argument. Following and building upon Popper's epistemology he proposes that the whole idea of epistemic *justification* is abandoned.

> *Nothing gets justified.* Instead of following the critical rationalists in replacing philosophical *justification* by philosophical *description,* we may urge the philosophical *criticism* of standards as the main task of the philosopher. *Nothing gets justified; everything gets criticized.* Instead of positing infallible intellectual authorities to justify and guarantee positions, one may build a philosophical program for counteracting intellectual error (pp. 112-13).

This is 'pancritical rationalism', a position which is willing to hold positions, including this position itself, open to criticism, never cutting off argument by resorting to faith or irrational commitment to justify some belief.

But either what Bartley is proposing is a purely formal or semantic position, the substituting of 'criticism' for 'justification', or it is a substantive thesis in epistemology which is in fact at odds with the thrust of his avowed pancritical rationalism. For what he is in fact providing is a *proposal* about the formation of beliefs, whereas what he thinks that he is doing is providing an argument for the superiority of pancritical rationalism over the critical rationalism of Ayer or the classical rationalism of Descartes, both of which in his view are forms of fideism and both embroiled in the *tu quoque* argument.

One obvious objection to pancritical rationalism which he considers is that the whole idea of criticism itself presupposes standards, that it is logically impossible to criticize except by means of certain criteria.

Bartley has two answers to this difficulty. One answer is to say that the premisses which any pancritical rationalist uses for the purposes of criticism are those which are unproblematic at the time he criticizes, not those which are regarded as unproblematic in some final, logically-privileged sense.

> When one belief is subjected to criticism, many others, of course, have to be taken for granted – including those with which the criticism is being carried out. The latter are used as the basis of criticism not because they are themselves justified or beyond criticism, but because they are *unproblematical at present*. These are, *in that sense alone and during that time alone*, beyond criticism (p. 122).

The much-vaunted criticism thus turns out to be a temporally local affair, and as such it is open to anyone, rationalist or fideist, to claim that he is 'critical' in this sense, that he criticizes other positions in the light of the beliefs which he holds as being unproblematic at present. But could such beliefs, accepted for the time being, themselves be overturned? Of course they could, but they would be overturned, and could only be overturned, using arguments based upon premisses which are themselves only temporally local.

So there is nothing in the first of his answers in defence of pancritical rationalism which has the slightest impact upon the *tu quoque* argument. There is nothing to stop an irrationalist taking up a position corresponding to that which Bartley himself takes up as a pancritical rationalist. Such a person would not be a classical or critical fideist or irrationalist. He would be a 'pancritical irrationalist'. For all that Bartley has so far shown to the contrary such an irrationalist is a perfect possibility.

This brings us to Bartley's second answer:

Pancritical irrationalism is impossible because of the standards of criticism to which the pancritical rationalist appeals. The 'pancritical irrationalist' does not and could not appeal to such standards. Bartley lists the standards as – the check of logic, of sense observation, of scientific theory and of the problem (p. 127). The four checks reveal the superiority of pancritical rationalism because they enable it to avoid the *tu quoque* argument, but all forms of irrationalism fail to avoid it.

> Consequently, the *tu quoque* argument cannot be used at all against pancritical rationalism. Theologians have argued that not only to abandon allegiance to Christ, but even to subject that allegiance to criticism, is to forsake Christianity. But for the pancritical rationalist, continuing sub-

> jection to criticism of his allegiance to rationality is explicitly *part* of his rationalism (p. 120).

> To be sure, to abandon logic is to abandon rationality as surely as to abandon Christ is to abandon Christianity. The two positions differ, however, in that the rationalist can, from his own rationalist point of view, consider and be moved by criticisms of logic and of rationalism, whereas the Christian cannot, from his own Christian point of view, consider and be moved by criticisms of his Christian commitment (p. 134).

Let us take it that Bartley's account of Christianity is accurate. What about the supposed asymmetry between his own critical position and that of the irrationalist equivalent?

Perhaps such an asymmetry holds in the case of some kinds of irrationalism but certainly not in the case of all, and it only *appears* to hold because of an equivocation over 'criticism' in the course of the argument. Sometimes in the argument 'criticism' is used substantively, as in the critical application of the standards quoted above, the four checks. But at other times it is used in a purely formal sense, as criticism in the light of some standard or other.

Supposing Bartley to be correct in his characterization of Christianity, what the Christian theologians may be objecting to, in the first quotation given above, is the use of the four checks critically, not the use of any checks critically. But in a precisely parallel way, what the pancritical rationalist would accept is not *any* criticism, but criticism in terms of the four checks. And similarly, in the second of the two quotations given, the two positions only differ because Bartley is not comparing like with like. He says that the rationalist can consider and be moved by criticism of logic and rationalism. But suppose that someone said that logic was defective because it was at odds with the word of God. Would the critical rationalist be moved by such a criticism? Presumably not. Again, Bartley says that the Christian theologian cannot, from his own Christian point of view, consider and be moved by criticisms of his Christian commitment. But why not? Granted, the Christian (as described by Bartley) cannot be moved by any criticisms based upon the four checks, but then critical rationalism cannot be moved by criticism based upon Christianity. And if critical rationalism *is* moved by criticism based upon the four checks then why could not the Christian theologian be moved by criticism which stems from the word of God? Provided that like is compared with like, the two cases are fully symmetrical, and the *tu quoque* argument remains.

What does this show? One thing that it shows is that Bartley is mistaken in thinking that pancritical rationalism has been shown to be superior by some objectively valid argument. There may be such an argument but he has not produced it. And his failure indicates that what he has done (as distinct from what he thinks he has done) is forcefully and eloquently to *recommend* one type of policy for the forming of beliefs. Perhaps it is possible to argue for such recommendations but the arguments will have to be much more tentative, subtle and provisional than the arguments Bartley has produced for showing the overall superiority of pancritical rationalism.

University of Liverpool,
P.O. Box 147, Liverpool L69 3BX

© PAUL HELM 1987

NON-SIMULTANEOUS CAUSATION

By DOUGLAS EHRING

SOME philosophers have argued that if *c* causes *e* then *c* will be simultaneous with *e*. One line of argument for this thesis attempts to show that other temporal relations, especially that of a cause wholly preceding its effect, are inconsistent with some generally accepted feature of causal relations. Michael Dummett, in *Truth and Other Enigmas*, points to two such features which are claimed to be incompatible with causation over a temporal gap: the fact that an effect's occurrence is certain given its cause, and the fact that the time of a cause's occurrence explains the time of occurrence of its effect (Michael Dummett, *Truth and Other Enigmas*, Duckworth, 1978, pp. 319–32). This second feature is also argued to be incompatible with non-simultaneous causation by Sarah Waterlow in her paper 'Backwards Causation and Continuing' (*Mind*, 83, 1974, pp. 372–87). In this paper, I will argue that neither Dummett nor Waterlow successfully demonstrate that non-simultaneous causation is impossible.

Dummett's first objection to the claim that causes may precede their effects is that there would then be no certainty that such a cause would bring its effect since something might intervene during the interval to hinder the operation of the cause (Dummett, p. 320). This objection may be interpreted in two different ways. On the first interpretation Dummett is arguing that whereas in the case of simultaneous causation there is no logical possibility of the non-occurrence of the effect given that the cause has occurred, there does exist such a logical possibility in the case of non-simultaneous causation. Understood in these terms the argument is not sound since, on most theories of causation, occurrence of the cause does *not* logically guarantee occurrence of the effect, even in the case of simultaneous causation.

On the second interpretation what is being argued is that if there were a temporal gap between cause and effect, it would be physically possible for the cause to occur without the effect, but that no such physical possibility exists in the case of simultaneous causation. This asymmetry, however, cannot be sustained on closer examination. First, causes need not be physically sufficient for their effects. Various convincing cases of probabilistic causation have been cited in the literature. In such probabilistic cases, the cause will not be physically sufficient for its effects even when cause and effect are simultaneous (let alone when they are non-simultaneous). Secondly, even if we grant that causes are physically sufficient for their effects, this asymmetry fails to hold. Let us suppose that c is physically sufficient for e if and only if a statement that c occurs, given a statement of various relevant laws, and a statement of initial conditions and meaning postulates, entails a statement that e occurs. In other words, if the laws of nature and circumstances are held constant, there is no logical possibility of the cause occurring without the effect. On this understanding of 'physically sufficient' the simultaneous and the non-simultaneous causes are symmetrical: if the laws and the circumstances are held constant, the occurrence of the cause guarantees the occurrence of the effect. Now of course there is a further step in Dummett's argument: in the non-simultaneous case, unlike the simultaneous case, the circumstances could be changed during the interval by way of an intervention which prevents the effect from occurring. This additional consideration, however, will not generate the asymmetry Dummett requires. First, it is certainly possible that the laws in both cases are 'iron laws', i.e. laws such that given the occurrence of c in the circumstance at a time t_1, e will occur at t_2 no matter how the circumstances change after t_1. Secondly, even without 'iron laws' the two cases are on a par with respect to intervention. In the simultaneous case it is possible to intervene *at the time* of the occurrence of the cause in order to prevent the effect by changing the circumstances.

Another, somewhat different way to make this point is as follows: from the fact that later intervention can prevent e after c has occurred, it does not follow that c is not sufficient for e. The sense in which a cause is sufficient for its effect, if it is sufficient, is that of being sufficient *in the circumstances*. Indeed, if non-simultaneous causation is possible, then the possibility of intervention after the occurrence of the cause does not show that c is not sufficient *in this sense* for the effect. Consider the following variation on the type of case mentioned above. The cause c occurs at t_1, the effect e at t_2, and the circumstance C relevant to the causal efficacy of c with respect to e occurs at t_2. In this case the possibility of preventing e from occurring by tampering with the circumstance C at t_2 does not show that c is not sufficient *in the circumstances* C for e. Or if the possibiliity of intervention shows as such in this case, c could not be considered sufficient for e when c and e are simultaneous, since tampering with the circumstances is also possible in that case.

Dummett's second objection to the possibility of a temporal gap between cause and effect is that the precise time of occurrence of the effect after the cause would, in such cases, be an 'irrational' matter: '. . . for if the effect does not take place immediately, what makes it come about when, eventually, it does? (ibid., p. 320). There are at least two interpretations of this objection. According to the first, there is some *a priori* basis for our expectation that effects not follow their causes, if such an expectation actually exists. Without further argument, however, it seems no more 'rational' that effects occur simultaneously with their causes than after their causes. Independently of experience, we have no more reason to expect one configuration of events than the other. On the second interpretation, the objection is that in the case of non-simultaneous causation there is an additional explanatory demand that does not exist for simultaneous causation. But again, as far as Dummett's argument goes, there is no more reason to think that one type of temporal relation between cause and effect deserves explanation but the other relation does not require explanation. Consider that there are an infinite number of temporal relations possible between cause and effect, with simultaneity being just one such possibility. In that sense we can ask of whatever temporal relation is realized, why *that* relation is realized rather than one of the other possible relations, even if the temporal relation instantiated is that of simultaneity.

This second objection, however, may perhaps be interpreted in yet a third way. For the purpose of developing a further version of the objection I will turn to Waterlow's paper in which a similar objection is presented. According to Waterlow, if temporal gaps existed between cause and effect, then either these gaps would be of some regular length or the length of the gap would vary without pattern. If the length varied without pattern, then 'a causal hypothesis could never be falsified, since there would be no time by which time the alleged effect must happen if it is to be counted an effect' (Waterlow, p. 378). In turn, she argues, without such a regular pattern, confirmation of causal hypotheses is not possible. For example, in the series $a_1 b_1 a_2 b_2 a_3 b_3$, where each successive member is later than the member before it, we would have no reason to treat b_1 as the effect of a_1, or b_2 or b_3 as the effect of a_1, for example. Matching up particular causes with their effects is impossible without some regular pattern of temporal relations. On the other hand, if the length of the gap is regular in any given type of case, we require some *explanation* for such regularities of this sort.

The first part of Waterlow's dilemma is that in a world in which the temporal gap between cause and effect varied without a regular pattern, we could not falsify or confirm causal hypotheses. This objection does not show that such a world is impossible, which is the question under discussion. Even if empirical determinations of causal connections could not be made, there still would be a fact of the matter as to whether such causal connections hold between events in that world. We could, for example, say what must be true of such events on various theories of causation, even if we could not ascertain whether or not such conditions are satisfied.

As for the second horn of the dilemma, Waterlow adds that accepting the regular length of the interval as a brute fact is

> tolerable if we are dealing with isolated cases; but the supposition we are considering is that the interval is a regular length in all cases of a given kind. And just as, with event regularities, we seem all but inescapably drawn, whether logically or psychologically, to invoke a causal nexus as explanation of the regularity, not merely as another name for it; so in this case we surely would be, where the regular companions are two events and a time gap (ibid., p. 379).

Given this need for causal explanation we would be tempted to treat the temporal gap itself as part of the cause or the effect. But the option of attributing causal efficacy to the passage of time *as such* is an unwelcome and unacceptable prospect.

But are we actually compelled to accept causal efficacy for time intervals if there exists a regular temporal gap between cause and effect of a certain kind? That there exists a regularity between x and y does not by itself require that we treat x and y as causally connected. There are various regularities where this conclusion is not drawn. Consider, for example, the regularity that holds between the lighting of matches and the *absence* of moisture. This absence, although in regular conjunction with the lighting, is not treated as part of the cause since causal relata are restricted to positive occurrences of a certain sort, e.g. events or event features. Similarly, if we found that the length of the temporal gap between cause and effect were regular in character, we would not treat the length of time as a cause or as an effect since the passage of time does not satisfy other relevant conditions for being a causal relatum. For example, on the view that causal relata are events, the passage of time will not qualify since it is not an event.

Furthermore, even if it were granted that in some cases in which the temporal gap between cause and effect followed a pattern there exists an unfulfillable demand for explanation that does not arise in the simultaneous case, we need not grant such an asymmetry in all cases of non-simultaneous causation. Consider the following type of example: c at t_1 is a cause of e at t_2, and d at t_2 is a cause of e. In turn, c's causal influence on e is direct, but depends upon the occurrence of d, i.e. d is part of the circumstances necessary for c's causal connection to e, but c does not cause d. Further, suppose that d is the final part of a set of sufficient causal conditions for e, and that without d, e would not have occurred. The objection under discussion asserts that the timing of a simultaneous effect relative to its simultaneous cause requires no explanation. Hence, in this case, that e occurs at t_2, the time at which d occurs, requires no explanation. In other words, e occurs precisely when the final member of the set of sufficient conditions occurs. In addition, the temporal gap between c and e is regular in the sense that e occurs at the point after c when d occurs, which we may assume is of predictable duration (e.g. c and d may be joint effects of some common cause which guarantees that d will occur at a certain regular interval after c). Hence, we can say either that we do not require an explanation of the interval between c and e or that we possess such an explanation. We would say the latter if we are emphasizing the role of d, since we have assumed that the timing of an effect relative to its simultaneous cause requires no explanation. Or, if we are emphasizing the question of when d itself comes about, we possess some straightforward explanation of that by way of d's causal history. In either case we are not faced with an unsatisfiable demand for explaining the interval between c and e.

Southern Methodist University,
Dallas, Texas 75275, U.S.A.

© DOUGLAS EHRING 1987

CAUSAL SUFFICIENCY RECONSIDERED

By STIG ALSTRUP RASMUSSEN

IN my paper, 'Ruben and Lewis on Causal Sufficiency' (ANALYSIS 42.3, October 1982, pp. 207-11), I attempted to defend David Lewis's account of causal dependence (D. Lewis, 'Causation', *Journal of Philosophy* 1973, pp. 556-67; reprinted in E. Sosa (ed.), *Causation and Conditionals*, Oxford 1975, pp. 180-91) against David-Hillel Ruben's contention that Lewis's account is deficient in a way that can be remedied only at the cost of making causal dependence symmetrical ('Lewis and the Problem of Causal Sufficiency', ANALYSIS 41.1, January 1981, pp. 38-41). At least, asymmetry would, according to Ruben, have to be ensured by supplementing the amended account in a manner unlikely to appeal to Lewis; e.g., by having recourse to the temporal precedence of cause over effect.

Lewis's acount has it that an event, e, is causally dependent on another event, c, just in case e would have occurred if c had, and e would not have occurred if c did not occur. Using Lewis's symbol for the counterfactual conditional operator and '$O(x)$' as an abbreviation for 'x occurs', Lewis's account is embodied in (1):

$$(1)\ C(e, c) \text{ iff } O(c) \mathrel{\Box\!\!\rightarrow} O(e) \text{ and } \sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e).$$

Ruben's worry was this. Lewis's idea seems to be that of capturing, in terms of his counterfactuals, the idea that e is causally dependent on c, just in case c is sufficient and necessary in the circumstances for e. But on Lewis's account of counterfactuals, the actual truth of '$O(c)$' and '$O(e)$' entail the truth of '$O(c) \mathrel{\Box\!\!\rightarrow} O(e)$' (D. Lewis, *Counterfactuals*, Oxford 1973, p. 26); hence it would seem that, in case c and e actually occur, nothing in Lewis's account captures the idea of causal sufficiency. Similarly, if c and e fail to occur, the counterfactual supposed to capture causal necessity holds vacuously. Ruben concludes that Lewis's account will have to be amended; and he suggests, in effect, that it be amended so as to read:

$$(2)\ C(e, c) \text{ iff } O(c) \mathrel{\Box\!\!\rightarrow} O(e) \text{ and } O(e) \mathrel{\Box\!\!\rightarrow} O(c) \text{ and } \sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e) \text{ and } \sim O(e) \mathrel{\Box\!\!\rightarrow} \sim O(c).$$

Now, however, causal dependence is a symmetric relation; and the only reasonable way to avoid this is, Ruben suggests, that of appealing to the temporal priority of c over e.

In my original attempt to extricate Lewis's account from this worry, I alleged that one may infer from '$O(c)$', '$O(e)$', and '$\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)$' the non-trivial truth of

$$(3)\ \sim O(c) \mathrel{\Box\!\!\rightarrow} (O(c) \mathrel{\Box\!\!\rightarrow} O(e));$$

and also infer from '$\sim O(c)$', '$\sim O(e)$', and '$O(c) \mathrel{\Box\!\!\rightarrow} O(e)$' the non-trivial truth of

$$(4)\ O(c) \mathrel{\Box\!\!\rightarrow} (\sim O(c) \mathrel{\Box\!\!\rightarrow} \sim O(e)).$$

Furthermore, I took the alleged derivability of (3) and (4) to ensure that Lewis's account could, after all, capture causal sufficiency and necessity in the troublesome cases.

This last move was, I think, quite reasonable; but, unfortunately, I was wrong about the derivability of (3) and (4): my derivations crucially relied on the mistaken belief that although Lewis's counterfactuals do not in general admit of contraposition, they *are* subject to that rule of inference in the special circumstances involved in my derivations. They are not, however; and, in consequence, my original defence misfires (I am grateful to I. L. Humberstone for pointing out my mistake). However, I still think that Ruben's suggested 'improvement' defeats the very purpose of Lewis's sort of account of causal dependence; and, in fact, there are several strategies available to the adherent of that account. Firstly, it deserves note that Lewis nowhere commits himself to the admissibility of paraphrasing (1) in terms of necessary and sufficient conditions in the circumstances. The option remains open to him to hold on to (1), grant Ruben's point that nothing in (1) captures causal sufficiency in case c and e actually occur, and reject the necessity and sufficiency idiom. Secondly, Lewis might hold on to (1), grant that the necessity and sufficiency idiom is admissible, but insist that there simply is no more to causal necessity and sufficiency than Ruben's point will allow. Thirdly, Lewis could hold on to (1), allow the paraphrase in terms of necessity and sufficiency, admit that Ruben's argument raises a genuine worry, and abandon his preferred account of counterfactuals. As Lewis points out (*Counterfactuals*, p. 29), his counterfactual logic can easily be weakened so as not to admit the inference to a counterfactual from the actual truth of its antecedent and consequent; and so altering the account of counterfactuals, if only by suitably restricting the set of admissible models in the pertinent cases, would clearly meet Ruben's point. Finally, there is the possibility of holding on to Lewis's preferred account of counterfactuals, admitting that the account of causal dependence should capture causal necessity and sufficiency in a way not vulnerable to trivialisation by the actual occurrence or non-occurrence of events, and granting that Ruben's worry indicates that (1) has to be improved upon in some suitable way. There is at least one fairly natural way of implementing the last strategy, if one is prepared to admit that had (3) and (4) been derivable in the way I previously thought them derivable, then Lewis's account would not have been vulnerable to Ruben's line of criticism: just add (3) and (4) to the right-hand side of (1). The overall account resulting on this move can do everything that Lewis's original account can do; but the account will now evidently be somewhat lacking in simplicity. If one is prepared to leave undefined 'causal dependence' for non-occurring events, a simpler account is however available: one could then replace (1) by

(5) $C(e, c)$ iff $\sim O(c) \Box\!\!\rightarrow (O(c) \Box\!\!\rightarrow O(e))$ and
$O(c) \Box\!\!\rightarrow (\sim O(c) \Box\!\!\rightarrow \sim O(e))$;

and in case a need is perceived for a general account, one may ensure that '$C(e, c)$' cannot be true unless '$O(c)$' and '$O(e)$' enjoy the same truth value by stipulating as follows:

(6) $C(e, c)$ iff $O(c) \equiv O(e)$ and $\sim O(c) \Box\!\!\rightarrow (O(c) \Box\!\!\rightarrow O(e))$ and
$O(c) \Box\!\!\rightarrow (\sim O(c) \Box\!\!\rightarrow \sim O(e))$.

University of Copenhagen,
Købmagergade 50,
1150 Copenhagen K, Denmark

© STIG ALSTRUP RASMUSSEN 1987

EXPLAINING CONTRASTIVE FACTS

By DAVID-HILLEL RUBEN

IT has become something of a fashion to regard what is being explained in an explanation as always, at least implictly, a contrastive fact (examples of this view include: Fred Dretske, 'Contrastive Facts' *Philosophical Review 81*, 1972, pp. 411-37; Alan Garfinkel, *Forms of Explanation*, Yale University Press, 1981, from which the Sutton story is borrowed; Bas van Fraassen, *The Scientific Image*, Oxford University Press, 1980). The 'old' view was that a typical object of explanation was an event, or an event under a description, or the fact that some event occurred (or perhaps, state instead of event). I do not here distinguish between these various non-contrastive possibilities.

An anecdote about the American bank robber, Willy Sutton, is intended to illustrate this contrastive view of explananda. When asked why he robbed banks, Sutton replied that it was because that is where the money was kept. The contrastive diagnosis of this anecdote is that the fact to be explained, that Sutton robs banks, is merely elliptical for some contrastive fact, and that there is more than one contrastive fact for which it might be elliptical. The inquirer was no doubt wishing for an explanation of Sutton's robbing banks rather than not robbing at all; Sutton replied with an explanation of why he robbed banks rather than institutions in which no money was kept.

I accept Dretske's view that a request for an explanation of why *e* occurred might raise different questions, depending on which word or words in the sentence are stressed or emphasised (Dretske, op. cit.). I also see no harm in speaking of contrastive facts, although it would be preferable to do so in some sense in which not all contingent facts were contrastive (we still want a contrast between contrastive and non-contrastive facts). However, on my view, contrastive facts are nothing more than a special sort of conjunction of non-contrastive facts. I take a broadly reductionist view of the matter.

What is involved in explaining why *e* occurred rather than *f*? Suppose according to some non-contrastive account I can explain why *e* occurred (suppose I can derive a sentence that says that *e* occurred from premises satisfying any conditions the reader wishes to impose). Since the explanation of *e*'s occurrence will typically leave it open whether or not *f* occurred, I could not, just in explaining *e*, plausibly claim to have explained why *e* occurred rather than *f*. For example, if I want to explain why I was in London rather than in Boston in the late evening of 7 January 1986, it is not sufficient to explain only why I was in London during that period. What is needed, in addition to the explanation of *e*'s occurring, is an explanation of how or why *e*'s occurring is inconsistent with *f*'s occurring (or, as I shall say, 'eclipsed *f*', where this is taken in a wider sense than just 'causally prevents'). I claim:

(1) A person explains why *e* rather than *f* occurred iff he explains why *e* occurred and how or why *e*'s occurring eclipsed *f*'s occurring.

Sometimes the additional information needed for this eclipsing is quite straightforward. For example, if I explain why I am in London on that late evening, I need only add the information which shows that it is physically impossible to be both in London and in Boston in the course of the same late evening, in order to have explained why I am in London rather than Boston during that time.

Explaining the conjunction of facts is to explain the contrastive. If I am asked to explain why I was in London rather than Boston, there will normally be a suggestion that I was presented with some choice specifically between those alternatives. But this is merely a matter of pragmatics, and the suggestion is cancellable. If I explain why I was in London, and I add the information that no one can be in two places at the same time, I will have explained willy-nilly why I was in London rather than anywhere else at the same time, whether or not that indefinitely large number of alternatives ever came before my mind. The eclipsing information eclipses whether or not these sorts of normal presuppositions about choices and alternatives hold.

A common assumption of the contrastive view, which was explicit in the diagnosis of Willy Sutton's little joke, is that these two facts are similarly contrastive: the fact that *e* occurred rather than not; the fact that *e* occurred rather than *f*. (1) shows us why these facts are essentially dissimilar. Suppose, once again, that I have explained, at least on some non-contrastive view, why *e* occurred. What more need I do to explain why *e* occurred rather than not? If we try to apply the idea of eclipsing to bring out what more is involved, we obtain: to explain why *e* occurred rather than not is to explain why *e* occurred and how or why *e*'s occurring eclipsed *e*'s failing to occur. As the discerning reader will have noticed, there is nothing more here for the eclipsing to explain. The only fact immediately relevant to how or why *e*'s occurring eclipsed *e*'s failing to occur is just the fact that *e* occurred. So

(2) A person explains why *e* occurred rather than not (rather than failed to occur) iff he explains why *e* occurred.

(1) shows how explaining a contrastive fact is to be analysed as explaining a certain conjunction of non-contrastive facts. (2) shows us that in the case of explaining why *e* occurred rather than not, we have not got a contrastive fact even in this reducible sense. 'Explaining why *e* occurred rather than not' is just a tedious

pleonasm for 'explaining why *e* occurred', which is to explain a non-conjunctive (and non-contrastive) fact. Indeed, if '*e*'s occurring rather than not' did name a contrastive fact, the very distinction between contrastive and non-contrastive facts would be at risk. Any apparently non-contrastive fact, whether or not it figured as the explanandum in some explanation, could be shown 'really' just to be the contrastive fact that that-fact-rather-than-not-that-fact.

How, on my view, shall we explain Sutton's little joke? I think that it is extremely common for us to believe that we have explained more than strictly we have explained, although in normal contexts it would be sheer pedantry to point this out. For example, if asked to explain why some criminal robs banks, I might produce an account of his deprived social background. Strictly speaking, my account can only explain, at most, why he entered the life of crime, but not specifically why he become a bank robber. When the priest asked Sutton why he robbed banks, what he really wanted an explanation of is a different fact: Sutton's engaging in a life of crime. An explanation of why Sutton robbed banks would *include* an explanation of why he chose bank robbing as the crime to engage in. Sutton took the priest literally at his word: he offered a *part* of the explanation of why he robbed banks, and indeed that part of the explanation in which the priest had no real interest. Of course, if Sutton had fully explained why he robbed banks, and explained why robbing banks was inconsistent in some way with robbing anything else, he would thereby have explained why he robbed banks rather than other institutions, since the latter explanandum is just reducible to the former.

The London School of Economics,
Houghton Street, London WC2A 2AE

© DAVID-HILLEL RUBEN 1987

ON AN ARGUMENT FOR INCOMPATIBILISM

By DAVID WIDERKER

CARL GINET ([1] pp. 182–3; [2] pp. 391–2) and Peter van Inwagen ([3] pp. 55–105) have recently defended a certain modal argument for the incompatibility of freedom and determinism. In this paper, I seek to examine this argument and to show that it is not sound. I conclude by suggesting another, in my view, more successful argument for incompatibilism.

At the heart of the Ginet–Inwagen argument for incompatibilism are the following two rules of inference which are assumed to be valid:

(α) $\Box p \vdash Np$

(β) $Np, N(p \supset q) \vdash Nq$

where 'Np' is short for 'p and no one has or ever had it within his power to make it the case that ~p', and 'p' and 'q' are sentence schemata ([2] p. 392; [3] pp. 94–6). (In van Inwagen's ([3] p. 94), 'Np' stands for 'p and no one ever has or ever had any choice about whether p', an expression which I take to be synonymous with 'Np' as defined above.) These two rules together with assumptions such as

(1) NP_0

(2) NL

(3) $\Box[(P_0 \,\&\, L) \supset P]$ by the truth of determinism

enable one to construct a deduction the conclusion of which is that no one ever has any choice about anything, e.g.,

(4) $\Box(P_0 \supset (L \supset P))$ By (3)

(5) $N(P_0 \supset (L \supset P))$ (4), (α)

(6) $N(L \supset P)$ (5), (1), (β)

(7) NP (6), (2), (β)

Here 'P_0' abbreviates a sentence 'about the remote past', e.g. 'about the distribution and momenta of atoms, and other particles in the inchoate presiderial nebulae', 'L' abbreviates the conjunction of all laws of nature, and 'P' schematizes any true sentence ([3] pp. 94–6).

Consider more closely the modens ponens rule for power necessity (β). Given the above explication of 'N', this rule may be stated in a more perspicuous way as follows:

$$(\gamma)\quad \begin{array}{l} p \,\&\, (x)(t) \sim P^*_{x,t}(\sim p) \\ \underline{(p \supset q) \,\&\, (x)(t)[\sim P^*_{x,t}(p \,\&\, \sim q)]} \\ q \,\&\, (x)(t) \sim P^*_{x,t}(\sim q) \end{array}$$

where 'x', and 't' range over agents and times, and '$P^*_{x,t}$' is short for 'it is within x's power at t to bring it about that'. My claim now is that (γ) may lead us from truth to falsity in situations involving the occurrence (non-occurrence) of uncaused, random events, or events associated with free human actions. Suppose that by destroying a bit of radium r before t9, Sam prevents the emission of a subatomic particle by r at t9. Suppose further that this is the only way by which Sam can make sure that r will not emit radiation at t9. Finally, suppose that Sam is the only sentient being that exists or ever existed. Let 'R' and 'S' stand for

R: A bit of radium r emits at t9 a subatomic particle.

S: Sam destroys r before t9.

Consider now the following argument:

(D1) 1. $\sim R \,\&\, (x)(t) \sim P^*_{x,t}(R)$
2. $\underline{(\sim R \supset S) \,\&\, (x)(t)[\sim P^*_{x,t}(\sim R \,\&\, \sim S)]}$
3. $S \,\&\, (x)(t) \sim P^*_{x,t}(\sim S)$

(D1) is an instance of (γ). However, its premisses appear to be true and its conclusion false. Here is another example. Suppose that Jim comes to visit Quincy at t8, whereupon Quincy by locking the front door prevents Jim from being able to leave thereafter. Suppose further that (a) this is the only way by which Quincy can ensure that Jim will not leave at t10, (b) though Quincy can prevent Jim from leaving at t10, he cannot prevent him from staying with him at t10, (c) for some reason, it was not within Jim's power before t8 to arrange things in such a way that he would be able to leave Quincy's house at t10, and finally, (d) Jim and Quincy are the only existing individuals. Let 'J' and 'Q' abbreviate

J: Jim leaves at t10

Q: Quincy closes the door at t8,

and consider the following argument

(D2) 1. $\sim J \,\&\, (x)(t) \sim P^*_{x,t}(J)$
2. $\underline{(\sim J \supset Q) \,\&\, (x)(t)[\sim P^*_{x,t}(\sim J \,\&\, \sim Q)]}$
3. $Q \,\&\, (x)(t) \sim P^*_{x,t}(\sim Q)$

Again, though (D2) is an instance of (γ), its premisses appear to be true and its conclusion false. Can Ginet and van Inwagen reply to these objections? It seems that their only way out would be to deny the truth of premiss 2 of (D1), or more precisely to defend the claim that

(GI) $(\exists x)(\exists t)\, P^*_{x,t}(\sim R \,\&\, \sim S)$

(Similarly also for (D2)). Let us examine (GI) more closely. How are we to understand the assertion made in (GI) that some agent x has it within his power to bring about the conjunctive states of affairs $\sim R \,\&\, \sim S$? One possibility would be to say that x has such a power only if

(X1) x has it within his power at t to cause it to be the case that $(\sim R \,\&\, \sim S)$,

The only other way to understand (GI) would be to say that it is within x's power at t to bring about that both $(\sim R \,\&\, \sim S)$ *indirectly*,[1] i.e. only if

---

[1] These two senses of power roughly correspond to what A. Plantinga has called 'the power strongly/weakly to actualize a state of affairs' ([6] pp. 172-3).

(X2) there is an act-type A such that
  (i) it is within x's power at t to perform A, and
  (ii) if x were to perform A, then it would be the case that ($\sim$R & $\sim$S).

However, in the situation under consideration neither of these two conditions is satisfied. Notice also that truth of (X2) cannot be defended by the claim that the counterfactual '$\sim S > \sim R$' i.e.

(8) If Sam had not destroyed r before t9, r would not have emitted a subatomic particle at t9,

turns out to be true on the Stalnaker-Lewis semantics for counterfactuals. For the fact that this semantics yields such a result is considered one of its serious drawbacks ([4] pp. 4-5; [8] pp. 271-3). What one is justified in asserting, given that R describes a random event, is not (8) but

(9) If Sam had not destroyed r before t9, it might have been the case that r would not have emitted a subatomic particle at t9.

I submit, then, that the Ginet-Inwagen argument for the incompatibility of freedom with determinism is fallacious. Can we suggest another argument for this conclusion?

In his well known article 'Fatalism' ([7] p. 58), Richard Taylor, in the course of arguing for the truth of fatalism, has made use of the following principle:

(T) No agent can perform any given action if there is lacking at the same or any other time some condition or state of affairs necessary for the occurrence of that act.

Although this principle by Taylor has been shown to be fallacious ([5] pp. 744-6), it can be easily revised to yield a correct principle, i.e.

> No agent can bring about a state of affairs if there is lacking at the same or any other time some condition logically (or causally) necessary for the obtaining of that state of affairs, provided that the agent does not have the power to bring about that missing condition.[2]

or to put it in terms of the notion of a sufficient condition

(TW) No agent can prevent the obtaining of a state of affairs for which there exists a logically (or causally) sufficient condition, provided that he cannot prevent that condition from obtaining, i.e.

---

[2] McCall [5] argues that Taylor's (T) holds only for past necessary conditions. The condition suggested in the text is more general.

$$[p \,\&\, \Box(p \supset q) \,\&\, {\sim}P^*_{x,t}({\sim}p)] \supset {\sim}P^*_{x,t}({\sim}q)$$

(for any agent x and any time t.)[3]

Notice that (TW) is not vulnerable to the counterexamples to (β) stated above. It is also, I think, a more intuitive principle than (β). Finally, with its aid we can formulate a sounder argument for the incompatibility of freedom and determinism, i.e.

(W) 1. $P_0 \,\&\, L$
2. $\Box[(P_0 \,\&\, L) \supset P]$
3. $\underline{(x)(t)[{\sim}P^*_{x,t}\,{\sim}(P_0 \,\&\, L)]}$
4. $(x)(t)\,{\sim}P^*_{x,t}({\sim}P)$ by (TW)

Those who find the operator 'N( )' convenient may state the above argument more economically thus

1. $N(P_0 \,\&\, L)$
2. $\underline{\Box[(P_0 \,\&\, L) \supset P]}$
3. NP

In this case, they would have to replace the modus ponens rule for power necessity (β) by

(β') $Np, \Box(p \supset q) \vdash Nq$[4]

Bar-Ilan University,
52100 Ramat-Gan, Israel

© DAVID WIDERKER 1987

REFERENCES

[1] C. Ginet, 'The Conditional Analysis of Freedom', in P. van Inwagen, ed., *Time and Cause*, D. Reidel, 1980, pp. 171-86.
[2] C. Ginet, 'In Defense of Incompatibilism', *Philosophical Studies*, 44 (1983), 391-400.
[3] P. van Inwagen, *An Essay on Free Will*, Oxford University Press, 1983.
[4] F. Jackson, 'A Causal Theory of Counterfactuals', *Australasian Journal of Philosophy*, 55 (1977), 3-21.
[5] S. McCall, 'Review of Steven Cahn's Fate, Logic and Time', *The Journal of Philosophy*, 65 (1968), 744-6.
[6] A. Plantinga, *The Nature of Necessity*, Oxford University Press, 1974.
[7] R. Taylor, 'Fatalism', *The Philosophical Review*, 71 (1962), 56-66.
[8] P. Tichy, 'A Counterexample to the Stalnaker-Lewis Analysis of Counterfactuals', *Philosophical Studies*, 29 (1976), 271-3.

---

[3] Carl Ginet suggested a similar principle, in the light of which I came to see the relevance of Taylor's (T).

[4] I would like to thank Carl Ginet, Jonathan Stavi, Harry Friedman, Dale Gottlieb, and Charlotte Katzoff for some excellent discussions on the Ginet-Inwagen argument for incompatibilism.

WILLIAMS' ARGUMENT AGAINST EXTERNAL REASONS

By BRAD HOOKER

BERNARD WILLIAMS, in his article 'Internal and External Reasons' (reprinted in his collection *Moral Luck*), attacks the view that there can be external reasons for action – that is, reasons that are not relative to the agent's 'subjective motivational set' (p. 102). Williams wants to be fairly permissive about what goes into this subjective motivational set: he mentions not only desires but also 'dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent' (p. 105). Williams' argument against the possibility of external reasons might now be reconstructed roughly as follows.

(1) Given that one had made oneself aware of the relevant empirical facts, including (perhaps through imagination) facts about what some alternative outcome would be like, one would be engaging in rational practical deliberation if one were (a) ascertaining what way of satisfying some element in one's subjective motivational set would be best in the light of the other elements in the set, (b) deciding which among conflicting elements in one's subjective motivational set one attaches most weight to, or (c) 'finding constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment'. (This account of rational practical deliberation can be distilled from pp. 103-5.) Thus rational practical deliberation in each of its forms has as its starting point the subjective motivational set had by the agent prior to this deliberation.

(2) A reason is *internal* just if it can be reached by rational practical deliberation which starts from the agent's antecedent subjective motivational set.

(3) From (1) and (2) it follows that any reasons the agent arrives at *by rational practical deliberation* will be internal reasons. (That is, if I could, by means of deliberating in any of the ways mentioned in (1), arrive at a new belief about what there is reason for me to do, then there was an internal reason for me to do the act in question to begin with (p. 109).)

(4) What there is reason for me to do is determined by what I would, if I deliberated rationally (and knew all relevant empirical facts), find there is reason for me to do.

(5) From (3) and (4) it follows that there are no such things as external reasons.

Not surprisingly, Williams anticipates the reply his opponent (whom he calls 'the external reasons theorist') is most likely to try:

> It might be said that the force of an external reason statement can be explained in the following way. Such a statement implies that a rational agent would be motivated to act appropriately, and it can carry this implication, because a rational agent is precisely one who has a general disposition in his *S* [subjective motivational set] to do what (he believes) there is reason for him to do. So when he comes to believe that there is reason for him to ϕ, he is motivated to ϕ, even though, before, he neither had a motive to ϕ, nor any motive related to ϕ-ing in one of the ways considered in the account of deliberation (p. 109).

But Williams complains that this explanation employs a belief the content of which is in question. He then asks, rhetorically,

> *What* is it that one comes to believe when he comes to believe that there is a reason for him to ϕ, if it is not the proposition, or something that entails the proposition, that if he deliberated rationally, he would be motivated to act appropriately? (ibid.)

Though at first this rhetorical question seems to be a witty refutation of the external reasons theory, on further reflection it seems to be either question-begging or impotent. Which fault it has depends on how 'deliberated rationally' is meant. It is question-begging if Williams means by 'rational deliberation' his own conception of rational deliberation, outlined in (1) above. The external reasons theorist will not only reject Williams' claim that external reasons are impossible but also be dissatisfied with Williams' conception of rational practical deliberation – in contrast to Williams, the external theorist is likely to think that (at least some) rational deliberation about reasons for action starts not from the agent's own subjective present motivations, but from some objective ('external') values or requirements, fixed independently of the agent's present motivations. So, the external reasons theorist will deny that the content of the proposition that there is a reason for one to ϕ is to be captured by the proposition that, if one deliberated rationally according Williams' conception of rational deliberation, one would be motivated to ϕ.

But now suppose Williams doesn't want to beg any questions by his reference to rational practical deliberation. Suppose, that is, that he wants the reference to be taken in a neutral way which does not rule out the external reasons theorist's own conception of rational practical deliberation. In this case, the external reasons theorist will be able to accept that the proposition that one has a reason to ϕ entails the proposition that, if one were rational and deliberated rationally, one would be motivated to ϕ. But now Williams' rhetorical question is impotent. For accepting this entailment does not undermine the possibility of external reasons.

Williams is aware of the point that his argument against the possibility of external reasons turns on his account of practical deliberation (p. 110). But what we need is an answer to the question this point raises: what is wrong with those views which hold that (at least some) rational practical deliberation starts from something external to the agent's present subjective motivational set?

The dispute between Williams and the external reasons theorist is ultimately over the starting points of practical deliberation. Obviously I have not tried to show that the external reasons theorist is, as against Williams, right about this. What I have argued is that Williams does not show, at least in the article in question, that there cannot be external reasons.[1]

Virginia Commonwealth University,
Box 2025, Richmond, Virginia, U.S.A.

© BRAD HOOKER 1987

---

[1] The central idea in this article is due to Robert Gay, to whom I am naturally very grateful. No doubt if he had written up his idea into an article, the result would have been much superior to mine. I also want to thank Madison Powers and Roger Crisp for helpful comments on an earlier draft.

NOT A COUNTEREXAMPLE TO MODUS PONENS

By E. J. LOWE

IN 'A Counterexample to Modus Ponens' (*The Journal of Philosophy* LXXXII, 9, September 1985, pp. 462-71), Vann McGee presents the following alleged counterexample to *modus ponens* (he also presents two others which, however, are constructed in essentially the same way):

> Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the polls results believed, with good reason:
>
> If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.
>
> A Republican will win the election.
>
> Yet they did not have reason to believe
>
> If it's not Reagan who wins, it will be Anderson. (p. 462)

The example fails to serve McGee's purpose because, I believe, it does not exhibit a genuine application of *modus ponens* – a rule which, in McGee's own words, 'tells us that from an indicative conditional $\ulcorner$If ϕ then $\psi\urcorner$, together with its antecedent ϕ, one can infer ψ' (ibid.). McGee clearly supposes that his example supplies premises of the form $\ulcorner\phi \Rightarrow (\psi \Rightarrow \theta)\urcorner$ and ϕ, and a conclusion of the form $\ulcorner\psi \Rightarrow \theta\urcorner$, where '$\Rightarrow$' stands for a conditionship relation discernible in the English indicative conditional which, allegedly, is intermediate in strength between strict implication and material implication (see ibid., p. 465). My view, however, is that, while it may be allowed that many English indicative conditionals require to be interpreted as involving a conditionship relation stronger than that of material implication (and which I shall continue to represent by '$\Rightarrow$'), in fact the first premise of McGee's example has the form $\ulcorner\phi \Rightarrow (\psi \supset \theta)\urcorner$ – so that the embedded conditional is a *material* conditional – whereas the conclusion is indeed of the form $\ulcorner\psi \Rightarrow \theta\urcorner$. This being so, though, the failure of the conclusion to follow from the premises cannot be taken to impugn the validity of *modus ponens*, which is not exemplified by the inference in question; the only sort of fallacy involved in passing from these premises to this conclusion is purely one of equivocation.

My point may be made as follows. Plausibly – or so I would urge – McGee's first premise is just equivalent to

> If a Republican wins the election, then either it will be Reagan who wins or it will be Anderson.

Observe that from this and McGee's second premise *modus ponens* would have us infer

> Either it will be Reagan who wins or it will be Anderson,

which it *is* reasonable to believe in the circumstances described, because it is reasonable to believe a disjunction when it is reasonable to believe one of its disjuncts. Of course, this disjunction is just equivalent to the material conditional whose antecedent is

> Reagan will not win the election

and whose consequent is

> Anderson will win it.

Clearly, however, it is not this material conditional that is expressed in McGee's example by the *non*-embedded sentence

> If it's not Reagan who wins, it will be Anderson

since we are invited to interpret this latter in a sense in which it expresses something *un*reasonable to believe in the circumstances described (unreasonable because of course in those circumstances it is reasonable to believe a conditional with the same antecedent but a contrary consequent). We should appreciate, however, that this very form of words *can* be interpreted in either of two ways: *either* as saying something of the form $\ulcorner\psi \Rightarrow \theta\urcorner$ (which interpretation is demanded of the non-embedded occurrence of the sentence in McGee's example), *or* as saying something of the form $\ulcorner\psi \supset \theta\urcorner$ – and that it is the latter interpretation which is demanded where the sentence in question appears embedded in McGee's first premise.

If it is wondered why the latter interpretation is not readily available for the *non*-embedded conditional in McGee's example, the answer is familiar and straightforward enough: no conversational point is normally served by *asserting* something of the form $\ulcorner\psi \supset \theta\urcorner$ where ψ and θ are reasonably believed to be false, any more than it is by asserting something of the equivalent form $\ulcorner{\sim}\psi \vee \theta\urcorner$. This indeed is why, although in the circumstances of McGee's example it would be *reasonable to believe* the disjunction

> Either it will be Reagan who wins the election or it will be Anderson,

it would not be conversationally appropriate to *assert* this.

Perhaps it will be felt that I have not done enough to *prove* that McGee's first premise really is of the form $\ulcorner\phi \Rightarrow (\psi \supset \theta)\urcorner$. But it is enough that this is not a patently implausible interpretation, as I have attempted to show by pointing to the plausible equivalence of that premise with a sentence of the form $\ulcorner\phi \Rightarrow ({\sim}\psi \vee \theta)\urcorner$. The burden of proof lies rather with McGee to show that his first premise genuinely *is* of the form $\ulcorner\phi \Rightarrow (\psi \Rightarrow \theta)\urcorner$, since it is he who is relying on this assumption in order to challenge a deeply rooted principle of deductive inference. As things stand, it is more reasonable to appeal to the *validity* of *modus ponens* to show that McGee has misinterpreted the form of one of the sentences he invokes in his example. That is to say, a reasonable degree of logical conservatism entitles us to see in McGee's example not a breakdown of *modus ponens* but rather a demonstration that the English indicative conditional is sometimes interpretable as a material conditional and sometimes not – something which, as it happens, I have argued for elsewhere on quite independent grounds (see my 'Indicative and Counterfactual Conditionals', ANALYSIS 39.3, June 1979, pp. 139-41). And, more generally, the lesson to be drawn is that logicians should be wary of jettisoning long-established principles of inference in response to the discovery of *apparent* counterexamples: almost always it is safer to conclude that one's recalcitrant linguistic intuitions are erroneous or confused and to try to set them to rights by laying bare distinctions to which one was blind.

One final point. Even if it is granted that McGee's first premise *actually* has the form $\ulcorner\phi \Rightarrow (\psi \supset \theta)\urcorner$, it might still be asked whether that very sentence could *also* be interpreted as expressing something of the form $\ulcorner\phi \Rightarrow (\psi \Rightarrow \theta)\urcorner$ and if so whether under such an interpretation it could be taken to express something reasonable to believe in the circumstances described in McGee's example. My answer is that I am far from sure that the kind of conditionship relation represented by '$\Rightarrow$' genuinely admits of this sort of embedding, but that to the limited extent that I can make sense of the suggestion that it does my reply to the second part of the question is 'No'. This is because the nearest I can come to allowing the sentence in question the proposed interpretation is to suppose that it might be understood as expressing in an elliptical way something that might more perspicuously be expressed by the following:

> If a Republican wins the election, then it will have been true to say that if it's not Reagan who wins, it will be Anderson.

And this, to the extent that I can make sense of it, expresses something which I consider it would have been *un*reasonable for those in the circumstances described in McGee's example to believe. (As may be gathered, one of the primary difficulties I see in the notion of embedding one '⇒' conditional within another in the way McGee envisages is the problem of accommodating their *tenses*, particularly where both conditionals are future tense ones: for straightforwardly embedding one such conditional within another as its consequent has the effect of transforming the tense of the embedded conditional from the future to the future future, so that the embedded conditional is not in fact equivalent to a non-embedded conditional of the same form – an effect which can only be overcome, and even then not entirely satisfactorily, by using the future perfect tense in the embedding conditional. But these are problems I cannot examine in detail here.)

University of Durham,
50 Old Elvet, Durham DH1 3HN

© E. J. Lowe 1987

PARFIT ON SELVES AND THEIR INTERESTS

By Eva Bodanszky

DEREK PARFIT begins *Reasons and Persons* (Oxford, 1984) by introducing 'the Self-interest Theory' of rationality. He says 'the *central claim*' of this theory is:

> (S1) For each person, there is one supremely rational ultimate aim: that his life go, for him, as well as possible (p. 4).

Parfit offers several arguments against the Self-interest Theory in the course of the book. Only one of these rests on his major metaphysical thesis about the nature of persons. In fact, this argument is one of the main links between his theory of persons and his theory of reasons. I wish to show that, if we take seriously the theory of persons, the argument fails.

The metaphysical thesis in question (Parfit calls it 'the Reductionist View') is that 'a person is not like a Cartesian Ego, a being whose existence must be all or nothing. A person is like a nation' (p. 275). According to the form of the thesis he defends, psychological connectedness and continuity replace personal identity – these relations 'are what matter' in the way personal identity is ordinarily thought to matter (pp. 275, 282, 307, 313).

If we accept the Reductionist View, Parfit claims, we must accept

> (C) My concern for my future may correspond to the degree of connectedness between me now and myself in the future. Connectedness is one of the two relations that give me reasons to be specially concerned about my own future. It can be rational to care less, when one of the grounds for caring will hold to a lesser degree. Since connectedness is nearly always weaker over long periods, I can rationally care less about my further future (p. 313).

But (C) contradicts a principle Parfit claims is 'central to the Self-interest Theory':

> *The Requirement of Equal Concern:* A rational person would be *equally* concerned about *all* the parts of his future (p. 313).

He concludes that we should reject the Self-interest Theory.

Parfit considers the reply that seems right to me – a version of the Self-interest Theory that leaves out the Requirement of Equal Concern:

> On the Revised Theory, a rational person's dominant concern should be his own future, but he may be less concerned about those parts of his future to which he is less closely connected (p. 317).

He rejects the revision as departing too far from the original to rescue it:

> This revision makes a big difference. It breaks the link between the Self-interest Theory and what is in one's own best interests. On the unrevised or *Classical Theory*, it is irrational for anyone to do what he believes will be worse for him. On the Revised Theory, this claim must be abandoned. If it is not irrational to care less about some parts of one's future, it may not be irrational to do what one believes will be worse for oneself. It may not be irrational to act, knowingly, against one's own self-interest.
>
> ... we must abandon the Classical Theory's central claim. Consider acts of gross or great imprudence. For the sake of small pleasures in my youth, I cause it to be true that I shall suffer greatly in my old age. I might, for instance, start smoking when I am a boy. I know that I am likely to impose on myself a premature and painful death. I know that what I am doing is likely to be much worse for me. Since we must reject the Classical Theory, we cannot claim that all such acts are irrational (p. 317).

Recall that the central claim of the Self-interest Theory is (S1). Since (S1) refers to people and their lives, we should, presumably, interpret it in light of the theory that these are not 'all or nothing'.

Parfit proposes that, instead of talking about people and their lives, we talk about 'successive selves': 'On my proposed way of talking, we use "I", and the other pronouns, to refer only to the parts of our lives to which, when speaking, we have the strongest psychological connections' (p. 304). He warns that this way of talking is suited only to cases 'where there is some sharp discontinuity, marking the boundary between two selves' (p. 306), but says also, 'Since this connectedness is a matter of degree, the drawing of these distinctions can be left to the choice of the speaker, and be allowed to vary from context to context' (p. 304). In the context of the Self-interest Theory, it seems appropriate to distinguish successive selves wherever the connectedness between two bits of life drops sufficiently to justify a difference of concern about their interests.

I think we may suppose that, for any pair of selves, s_i and s_j, the extent to which they are connected can be represented by a number between 1 and 0, $C(s_i s_j)$. 1 represents maximal connectedness, so that, normally, $C(s_i s_j) = 1$ only if $s_i = s_j$. If s_i is not at all connected with s_j, $C(s_i s_j) = 0$, and intermediate degrees of connection get intermediate numbers. Let '$V(s_i)$' represent the extent to which things go well for s_i – the better they go, the higher the number. Each non-zero product $C(s_i s_j)V(s_j)$ will then represent the extent to which things go well for one of s_i's successive selves, weighted according to the connection between that self and s_i.

I suggest that one natural rendition of (S1) in this reductionist way of talking is:

> (RS1) For each self, s, there is one supremely rational ultimate aim: that $\Sigma_{s_i}C(ss_i)\,V(s_i)$ be maximal.

The 'classical' self-interest theory follows from (RS1) together with the classical assumption that, for each self, there is just one self connected with it.

(RS1) does license acts ordinarily regarded as imprudent, such as smoking when a boy, when the future suffering is sufficiently mitigated by decreased connection. But Parfit is not entitled to describe this as the consequence that 'it may not be irrational to do what one believes will be worse for oneself'. On his account, the present self will not be around by the time things get worse, and the self that will then be its successor is not fully connected with it.

Nor is (RS1) open to Parfit's two minor objections. He says it is a weakness of the Revised Theory that it introduces a 'discount rate', hence needs to 'explain what makes a discount rate too steep' (p. 318). (RS1) discounts the interests of future selves just in proportion to their lack of connection with the present self.

He claims also that the Revised Theory is not a version of the Self-interest Theory but a version of the Critical Present-aim Theory (p. 317). On the Critical Present-aim Theory, 'each person has most reason to achieve those of his present aims that are not intrinsically irrational' (p. 92). (RS1) says nothing about anyone's actual present aims, in fact, nothing about what makes things go well for a self. Parfit discusses, but declines to decide among, three 'theories of self-interest': the 'Hedonistic Theory', the 'Desire-fulfilment Theory', and the 'Objective List Theory' (p. 4). It seems to be the Desire-fulfilment version of (RS1) that is in most danger of approaching the Critical Present-aim Theory, but even that doesn't quite arrive – it counts the desires of future selves less heavily than it counts present desires, but it does count them. Parfit earlier claims that *every* theory of rationality is a version of the Critical Present-aim Theory – given suitable interpretation of 'intrinsically irrational', the actual present aims just drop out (p. 194). I'm not convinced about this, but if it were true, it would be misleading to suggest that being a version of the Critical Present-aim Theory tells against being a version of the Self-interest Theory.

I think the point of the Self-interest Theory has not been lost in the translation. In particular, the revised self-interest theory can be contrasted with other theories of rationality in much the way the old one can. The contrasts are not as sharp as before, since we are supposing that people might be, so to speak, fuzzy around the edges, but they remain recognizable contrasts. The self-interest theory remains a rival to universalistic theories of rationality as well as to (standard) present-aim theories.[1]

University of Massachusetts,
352 Bartlett Hall, Amherst, Massachusetts 01003, U.S.A.

© EVA BODANSZKY 1987

---

[1] I am grateful to Fred Feldman for comments on an earlier draft of this paper.

SURVIVAL AND TRIVIAL FACTS

By DOUGLAS EHRING

ACCORDING to Derek Parfit, what makes the difference between survival and death cannot be some trivial fact (Derek Parfit, *Reasons and Persons*, Oxford, Clarendon Press, 1984, p. 271). What makes for survival is psychological connectedness and/or psychological continuity, Relation R. In this paper, I argue that given this non-triviality requirement, Relation R can be shown not to be what is important to survival.

In case of the ordinary survival of a person, what matters is psychological connectedness and/or continuity, Relation R, according to Parfit. Relation R is constituted by various component relations. For the sake of discussion, I will focus on only one – memory, or more precisely, quasi-memory. X quasi-remembers having an experience just in case X seems to remember having an experience, someone Y did have this experience, and X's apparent memory is causally dependent in the right kind of way on that experience. Parfit adopts a 'Wide' view of psychological connectedness according to which 'causally dependent in the right way' is 'causally dependent in any way' (ibid. p. 262). X has a direct memory connection with Y if X can quasi-remember some of the experiences Y had. There is *strong* connectedness of quasi-memory if, over each day, the number of direct quasi-memory connections is at least half the number in most actual lives (ibid. p. 222). There is continuity of quasi-memory between X and Y if an overlapping chain of strong quasi-memory connections exists between X and Y.

X and Y are the same person just in case X is R-related to Y and there is no different person who is also R-related to X (ibid. p. 267). Relation R generally coincides with personal identity since under normal conditions this second conjunct, requiring that Y be unique in being R-related to X, is satisfied. However, it is possible for different people to be R-related to the same person X. In such cases, Relation R will not coincide with personal identity. Nevertheless, what is important in ordinary cases of survival, Relation R, would be realized in these possible cases. As Parfit puts it, Relation R 'is what matters even when, as in a case where one person is R-related to two other people, Relation R does not provide personal identity' (ibid. p. 217).

In defending this view of what matters in ordinary survival, Parfit argues that his analysis satisfies an analogue to the following plausible principle suggested by Bernard Williams:

> Since personal identity has great significance, whether identity holds cannot depend on a trivial fact (ibid. p. 267).

According to Parfit, no plausible criterion of personal identity will satisfy this requirement. An analogue 'non-triviality' principle, which does not assume that survival must involve personal identity, can be stated as follows:

> Since survival has great significance, whether survival holds cannot depend on a trivial fact.

This principle implies that if there is only a trivial factual difference between two cases, there should not be a difference concerning the facts of survival. If, for example, S survives in one such case, then it should *not* be that S dies in any case which is factually only trivially different. Hence, if Relation R is what matters in survival, Relation R should satisfy this principle. Indeed, Parfit claims as much: 'Unlike identity, this relation cannot fail to hold because of a trivial difference in the facts. If this relation fails to hold, there is a deep difference in the facts' (ibid. p. 271).

The following sequence of cases will show, however, that Relation R does not satisfy this non-triviality condition.

Case 1: At some point in the distant future, medical technology makes possible the encoding on tape of information detailing the structure of a person's brain. As a result of these advances, it has become common in old age to undergo a transfer of one's memories and personality traits to a new, healthy body by way of this encoding process. After the relevant information is recorded, the original body is destroyed and a new brain in a new body is restructured on the basis of the information on the tape. If we call the person with the old body A and the person with the new body, after the restructuring, B, then we can suppose that B seems to remember having experiences which A had and that these apparent memories are causally dependent in some way on the experiences of A. There exists, then, psychological connectedness and/or continuity between A and B, i.e. Relation R. The outcome of the procedure, on Parfit's view, should be viewed by A, in advance, as being as good as ordinary survival since that which is important in ordinary survival, Relation R, is realized.

Case 2: In this case, an unlikely causal sequence occurs. After encoding information on a tape from A and after A's destruction, the tape is accidentally dropped onto another encoding machine. As a result the tape is damaged beyond repair. However, the tape's hitting the machine causes the machine to malfunction with the highly improbable outcome that this machine produces another tape exactly similar to the original. Although the original tape plays a causal role in the production of the new tape, the latter tape is not produced by *copying* the first tape. Technicians transfer the information on this new tape to a new brain. The resulting person B will be R-related to the original person A since B will have apparent memories of experiences A had and these apparent memories will be causally dependent in *some* way of those experiences. Given that Parfit allows *any* causal route from A's experiences to the apparent memories of B, he must admit that A and B are R-related. Hence, in this case, A ought to view, in advance, the outcome of this sequence as being as good as ordinary survival.

Case 3: In this final case, just as in case (2), the tape which contains the information extracted from A drops onto another encoding machine and is damaged beyond repair. However, unlike case (2) the tape's hitting the machine does not cause it to produce an exactly similar tape. Nevertheless, an exactly similar tape is produced at just that moment by this encoding machine as a result of an unrelated accident (e.g. by way of a short-circuit not caused by the tape falling onto the machine). Technicians transfer the information from this exactly similar tape to a new brain in a new body. The resulting person B has apparent memories of A's experiences, but these apparent memories are *not* causally dependent in any way on the experiences of A. Hence, even on Parfit's liberal account of the R Relation which countenances any causal route, A and B are not R-related (assuming no other component relation of the R Relation is realized). Hence, A ought to view, in advance, the outcome of this sequence as being as bad as ordinary death.

Relation R holds between A and B in cases (1) and (2), but not in case (3). However, the factual difference between case (2) and (3) is trivial. (2) and (3) differ only with respect to the 'accidental' cause of the duplicate tape. Although there is a causal chain running from A's experiences to B's apparent memories in (2), this causal chain is rather haphazard. Case (3) is generated from (2) by slightly modifying the nature of the 'accidental' character of the sequence. Instead of the dropping of the original tape accidentally and improbably resulting in the duplicate tape, the latter is the result of another independent, but equally improbable causal sequence. If A would be willing to view case (2) in advance with optimism, it would then seem irrational for A to view case (3) with profound pessimism concerning his fate given this relatively trivial difference in the causal facts. However, if Relation R is what is important in survival, just such a change in outlook would be appropriate. But one wants to say, given the non-triviality condition, that if A survives the accidental sequence in (2), he must also survive the slightly more accidental sequence in (3). Hence, since the factual difference between (2) and (3) is trivial, but Relation R holds in the former but not in the latter, Relation R does not satisfy the non-triviality requirement.

One might object that the difference between cases (2) and (3) is *not* trivial since the difference involves a break in the causal chain from A's experiences to B's apparent memories. This response, however, can either be seen to beg the question or to be too strong. This reply begs the question if it involves simply *assuming* that the only *important* differences in causal history are those which involve a failure in Relation R, other differences in causal history being trivial. On the other hand, if this assumption is not made, and it is left open as to what differences in causal history might be important, then this objection can be shown to be too strong. If the question of what are non-trivial differences in causal history is left open, then if the difference between (2) and (3) is judged profound, so must the difference between (1) and (2). In fact, the difference in causal history between (1) and (2) would seem to be greater than that between (2) and (3). The shift from case (1) to (2) involves a shift from what might be called a 'non-accidental' causal dependene of B's apparent memories on A's experiences to an 'accidental' causal dependence. The move from (2) to (3), however, preserves the 'accidental' character of the causal history of B's apparent memories, while eliminating the relevant causal dependence. If the latter difference in fact is held to be profound, it is reasonable to treat the former difference as perhaps even more profound. In other words, if the introduction of a significantly different type of causal same may be said of the existence of tribal deities, elves, hobgoblins, leprechauns or indeed of the existence of any preternatural being one may care to invent. But a theory which requires us to suspend judgment about the existence of such entities must be open to serious question. Surely the reasonable attitude is that they are nothing more than the products of the human imagination.

A second difficulty in Morris's position is that it would lead to almost total scepticism about human knowledge. We are clearly in a particularly poor epistemic position with regard to the existence of Descartes's evil demon, since if he did exist, he would ensure that evidence for that fact is concealed from us. But if this implies that we do not know whether this evil demon exists or not, then we do not know whether or not we are being constantly deceived. And this is equivalent to saying that we know nothing except those things about which we are incapable of being mistaken.

Where has Morris gone wrong? The answer, I believe, is that he has conflated two different sorts of existence claims – claims that something exists simply (what I shall call 'pure existence claims', e.g. 'God exists') and claims that something exists at a certain time and place (what I shall call 'qualified existence claims', e.g. 'There is a fly in my soup'). The notion of a good epistemic position has application with regard to the second category but not with regard to the first. For Morris

> being in a good epistemic position relative to a proposition is being in such a position that should there be any positive epistemic considerations for the truth of the proposition, one would have them or most likely have them (ibid. p. 222).

Thus if someone says that there is a snake in my room, I can put myself in a good epistemic position for assessing that claim by searching the place. But there is no way in which I can put myself in a good epistemic position with regard to a pure existence claim, such as 'Unicorns exist', since I cannot search the entire universe. It is true that I can sometimes know a pure existence claim to be true, but this establishes no more than that I can have sufficient evidence in favour of the claim to warrant acceptance of it. And this is obviously very different from saying that should there be evidence in support of the claim, I would possess, or be likely to possess, it.

Morris appears to be confused on this point, for he believes that if we could prove or disprove a metaphysical claim, such as that God exists, then we would be in a good epistemic position relative to that claim (ibid. p. 233). But the fact that I can prove or disprove the existence of God does not imply that I possess, or am likely to possess, all the evidence relative to his existence. Indeed, assuming there is such evidence, it would be difficult for anyone other than an omniscient being to have access to all of it. It is of course paradoxical to assert that I know a pure existence claim, such as 'Tigers exist', to be true and yet am not in a good epistemic position for assessing it. But from the fact that I know that tigers exist it clearly does not follow that I possess, or am likely to possess, all the evidence relative to their existence. No doubt it is theoretically possible for someone to possess all the evidence concerning the existence of tigers. But one could never be in a position to assert that that is all the evidence there is, since there is no guarantee that tigers do not also exist in other, utterly remote, parts of the physical universe. (If the word 'tiger' is understood to refer exclusively to a terrestrial species, then 'Tigers exist' is no longer a pure existence claim.) The notion of being in a good epistemic position would appear to be unusable therefore so far as pure existence claims are concerned.

However, it might be argued that this difficulty could be overcome if Morris's account of what is involved in being in a good epistemic position relative to a proposition were to be emended. Thus it could be understood to mean 'being in such a position that should there be positive evidence in favour of the proposition, one would most likely possess *some* of it'. But this will not do, since one can possess some evidence in favour of a proposition while remaining in a poor position for assessing its truth. Thus there is some evidence which suggests that there is intelligent life elsewhere in the physical universe, but we are not, at present, in a good position for assessing the truth of that claim. Alternatively, it might be understood to mean 'being in such a position that should there by any positive evidence in favour of the proposition, one would possess *enough to warrant acceptance of it*'. But this will not do either, as is clear from the following example. Suppose I receive a letter from someone claiming to be a disgruntled student, which states that there is a bomb in my office. I go to the police, who search the place thoroughly, but find nothing. I am not now in a good epistemic position relative to the proposition 'There is a bomb in my office', since though there is some positive evidence in favour of it, there is not enough to warrant acceptance of it. And this means, according to Morris, that no amount of merely negative evidence will warrant rejection of it, so that I must remain agnostic about the existence of the bomb no matter how thoroughly my office has been searched. But this conclusion is irrational; the absence of positive evidence in such a situation is as decisive a consideration as one could want.

As a final resort one might understand 'being in a good epistemic position relative to a proposition' to mean 'being in such a position that should there be enough evidence available to warrant acceptance of the proposition, one would be in possession of it'. This would avoid the difficulty just mentioned, but it would do nothing to alleviate the other difficulties in Morris's position, for since one could know that there was enough evidence available to warrant acceptance of a proposition only if one were in some sense in possession of that evidence, this new interpretation would imply that one could not be in a good epistemic position relative to a proposition unless one was justified in accepting it. This in turn would mean, according to Morris, that one should be an agnostic about all existence claims which one did not have good reason for believing to be true, so that it would be irrational to deny the existence of hobgoblins, Lilliputians, Descartes's evil demon etc.

Since it does not seem possible to emend Morris's account in any other way that is remotely plausible, I conclude that there are no good grounds for placing limits on the application of Ockham's razor to pure existence claims. In the absence of positive evidence for God's existence, therefore, one should be an atheist rather than an agnostic.

University College,
Cork, Ireland

© P. J. McGRATH 1987

THEODICY WITH A GOD OF LIMITED POWER: A REPLY TO McGRATH

By MICHAEL B. BURKE

P. J. McGRATH (in 'Evil and the Existence of a Finite God', ANALYSIS 46.1, January 1986, pp. 63-4) challenges the common assumption that the problem of evil could easily be solved if we were willing to conceive God as limited in power or goodness. So far as I can see, McGrath does succeed in showing that to conceive God as but imperfectly good would contribute little to a solution of the problem. McGrath does not succeed, I will argue, in showing that it would contribute little to conceive God as finite in power.

Here is McGrath's reasoning.

> Consider first what happens to the problem when God is conceived as infinitely good but not omnipotent. Mackie's thinking is presumably that evil can now be explained by saying that while God does not want evil to be present in the universe, he does not have sufficient power to prevent it. But the trouble with this is that some evils which formerly existed have been eliminated by human ingenuity. For example, smallpox has been wiped out through the use of vaccination. If man can get rid of an evil like smallpox, why could God have not done likewise? To say that he was unable to do so is to reduce his power to such an extent that it is difficult to see how he could be any longer regarded as divine. Certainly a being which lacks the power to do something which man is capable of doing could scarcely be a proper object of worship. (pp. 63–64)

If God would have eliminated all evils if he could have, why didn't he at least eliminate smallpox? Was he unable even to do that? McGrath overlooks the reply that God failed to eliminate smallpox not because such a feat was beyond his power, but because all of his vast but finite power was required for the elimination or prevention of *greater* evils. It could be suggested, to give just one example, that our galaxy is but one of hundreds of millions with intelligent life, that God can attend to 'only' a dozen galaxies at a time, and that ours is not among those most urgently in need of his attention.

Of course, the proposition that God would have eliminated smallpox if he could have is not a *logical implication* of the propositions that smallpox is an evil and that God would have eliminated *all evils* if he could have. To think it is would be to commit the fallacy of division.

Indiana University,
Indianapolis, Indiana 46202, U.S.A.

©MICHAEL B. BURKE 1987

GERALD VISION AND INDEXICALS

By JULIA COLTERJOHN *and* DUNCAN MACINTOSH

GERALD VISION ('I am here now', ANALYSIS 45.4, October 1985, pp. 198-9) offers supposed counterexamples to the 'common wisdom' (which he calls 'the indexical thesis'): (1) that the sentence (the referring terms of which he evidently presumes to be indexicals), 'I am here now' (hereafter 'S') cannot be said falsely; (2) that 'here' always refers to the vaguely bounded, present location of the utterer; and (3) that 'now' always refers to the time of utterance.

His first example: he points to the wrong place on a map while saying 'S'. Now, while we agree that the sentence here expresses a false proposition, we think its referring terms are being used *demonstratively*, not *indexically*.

The referents of both kinds of terms depend on the context of use. (We say 'use', rather than the more usual 'utterance', because indexicality can also figure as a property of writings, and because we shall have to distinguish between the times of *production* and *use* of an expression.) Thus, for example, in both indexical and demonstrative references the persons, times or places referred to depend on who is using the terms, when, and where, respectively.

Indeed, the referents of indexicals are identical to these. And the purely indexical use of the referring terms of 'S' guarantees that while the proposition it expresses is itself contingently true or false, no proposition *expressed or invoked that way* can be one that happens to be false. But since the use of demonstratives essentially involves one or another species of *demonstration* or ostension (pointing and the like), they can refer to things different from what their referents depend on, and given certain ostensions and contexts, the propositions expressed can be false ones. In Vision's example 'here' obviously functions *demonstratively*, a function more apparent in a sentence like, 'I am *there* now'. It is not a false *indexical* sentence. This also saves (2) above.

At best, Vision has exemplified a non-indexical use for terms often figuring as indexicals, refuting anyone who ever thought that certain (English) *terms* always had the referents needed for the sentences in which they figured to be true, or to have expressed true propositions. But surely this was only meant to hold when the relevant terms are *used* indexically. And though some referring terms have both indexical and demonstrative uses, English might as well have had separate words for each – say 'here' and 'teer', respectively. Thus, 'I am teer now' would be false in the example, but 'I am here now' must be true. Vision's case, then, depends entirely on the conflation of the two uses.

He further argues that 'S' can be false by referring to a time different from that of 'utterance'. His example: he leaves a note for his wife saying 'S', intending that she get it when she arrives home while he is upstairs asleep. But he goes out, forgetting the note, which apparently communicates a falsehood to her upon her arrival, two hours later.

Now we think there are two more plausible possibilities about what has happened here than Vision's conclusion against the common wisdom. (He rightly rejects the defence which holds that the note refers to the writer and the time and place of writing.)

(a) His wife's reading the note might be an extended case of *him using* the sentence (since *he* wrote it, left it, intended to communicate with it, and succeeded *when* she read it; obviously here the occasion of 'utterance' differs from the time of use by two hours). If it is indeed an indexical expression, its referents are the user (Vision), the time of usage (the time of reading), and the place of usage (where Vision then is). And wherever he is, 'S' used by him expresses a proposition true of him.

This saves the common wisdom, but doesn't explain his wife's realizing this is not just her husband ambushing her with philosophic truisms; she understands that he is *supposed* to be currently in the house, wherever he *happens* to be when she reads the note.

Thus a second proposal, one which does justice to the referents relevant to the achieved use of the expression: (b) perhaps the note is a 'proxy-finger', which can be used to refer to wherever it itself is, just as Vision can refer to any place charted by the map by pointing to the appropriate place on it. His wife, understanding the pragmatics of this kind of 'pointing' realizes the note refers to its present general location. 'I' and 'now' refer as indexicals to the user (Vision) and the time of use (when his wife reads the note), respectively. But the place of use (the house) is fixed by a convention of pointing which allows it to be picked out by Vision 'from a distance'. Since 'here' appears *demonstratively*, the sentence, though false, is *non*-indexical, and is still no counter-example to the common wisdom on indexicals.

(Our colleague William Blackburn has suggested that the sentence might be functioning indexically after all, 'here' referring to the location of the *speech act* that the note facilitates, which in this case is, arguably, the location of the house. While this interpretation would vindicate Vision's thesis, we don't think sense can be made of the house being the location of the speech act without invoking something like a convention of demonstration. Is the location of a speech act to be that of the sender or that of the receiver of a communication? It can be either. Suppose person one in Canada teletypes a message to person two in England, causing 'S' to appear on paper in England. Does the expression refer to Canada, and so say something true of person one, or to England, and so say something false of him? This depends, perhaps, on person one's referential intentions. And supposing he was intending to delude person two into thinking he (person one) was in England, wouldn't he have to invoke a convention of ostension? He is not (necessarily) to be taken to be where the message originates, but where it winds up. Somehow, he has to 'throw' the reference of the expression away from his present location. So call them what you will, there are clearly two different possible uses here, one where the reference is to the location of the speaker, another where the reference is to the location of a communicative artefact or of its recipient. The sentence cannot express a false proposition under the former use.)[1]

Dalhousie University,
Halifax, Nova Scotia B3H 3J5, Canada

© JULIA COLTERJOHN *and*
DUNCAN MACINTOSH 1987

---

[1] The authors are grateful to the Killam Trust of Dalhousie University for the Killam Special M.A. and the Killam Post-Doctoral fellowship support, respectively, which they enjoyed during the writing of this note.

HERE AND NOW

By PAUL SIMPSON

GERALD VISION in 'I am here now' (ANALYSIS 45.4, October 1985) takes exception to what he terms 'the indexical thesis'. The thesis he opposes would have the words of his title true whenever asserted and would imply that the sentence even frustrates attempts to lie with it.

His first false 'I am here now' is occasioned by its utterer misreading a map and pointing as the words are said at a part of the map that does not represent the utterer's environs. Vision holds that his example shows further that the denotation of 'here' need not be the vaguely bounded present location of its utterer. As his example apparently hinges on the ambivalence of reference of the term 'here' applied to part of a map, he has unnecessarily complicated the issue. For all parties should readily enough admit that, once an utterer points *away from* his or her own person to gain a reference to a public object, there will be no question of 'here' being used indexically or self-referentially. Thus one need not be pointing at part of a map in order to say 'I am here' falsely. One can point to anything else, a microscope slide or a table-top, and it is guaranteed that one will not be there – or as the gesturing speaker says, 'here'.

Of wider interest and perhaps undue concern is the fact noted by Wittgenstein (*Philosophical Investigations* I, section 185 and elsewhere) that what is picked out by pointing, and indeed our understanding of the entire convention, must remain under-determined. No amount of ostension or verbalizing will absolutely secure a common reference.

Vision intends 'now' in his later examples to refer, not to the time of utterance, but to the time of hearing or reading a message. He gains rather less mobility of reference for 'now' than dictionaries grant. Both *The Oxford English Dictionary* (1933) and *Webster*'s (1971) give (amongst others) the sense, 'at the time ... referred to', *O.E.D.* instancing 'The war was now practically concluded' (and thus permitting 'The war will now be practically concluded'). What Vision cannot do with his examples of people who leave messages for later inspection or audition, 'I am home now...', is to provide any compelling reason, other than to misinform, for retaining the present tense when the future is indicated. His utterers of 'I am home now ...' are either making predictions or they are not. If they are making predictions, they are using the wrong tense. If they are not making predictions, their later absences do not render their claims false. Once misinforming is allowed, his examples are once again needlessly artful. A telephone answering machine that always responds, 'I am home now', regardless of the original speaker's whereabouts, will serve that purpose.

An indexical thesis of the form opposed by Vision is very much a straw man. A departed soul or a time-traveller might succeed in transmitting audible messages to present-day people. Upon attempting to visit our world in person, the departed soul or time-traveller might fail, but thinking the mission to have succeeded announce, 'I am here now'. As the speaker has failed to arrive in person, we may speculate about the here and now being referred to, but at all events it is not our here and now. The 'here' and 'now' just used, usages also listed in dictionaries, clearly refer to matters that are in a sense public property and not the private preserves of their utterers. They are available for observation, measurement, and so forth, by an audience as much as by a speaker. They are hence detachable from utterances or acts of utterance and cannot be analysed completely, if at all, indexically.

University of Tasmania,
G.P.O. Box 252C,
Hobart 7001, Australia

© PAUL SIMPSON 1987

FOLEY ON CAUSATION AND RATIONALITY

By SCOTT STURGEON

IN 'Epistemic Luck and the Purely Epistemic', Richard Foley argues for the radical view that epistemic rationality is not bound by any sort of causal requirement (R. Foley, *American Philosophical Quarterly*, 1984, pp. 113-24). According to Foley, epistemic rationality places no restrictions on what sort of factors may be causally responsible for the production of a rational belief. Yet this is highly counter-intuitive. For we can easily imagine instances in which a person not only has good evidence to believe *p* but does believe *p*, and yet nonetheless is irrational precisely because he is caused to believe *p* for an irrelevant reason.

To counter these intuitions Foley claims that in such cases the irrationality of the belief involved can be accounted for in terms of a non-causal account of rational belief. For Foley, this account simply 'requires a person not only to have adequate evidence for his belief but also to appreciate how this evidence indicates his belief is true' (p. 115). But this too should seem puzzling. For it also seems easy to imagine situations in which a person has adequate evidence for *p*, recognizes it to be such evidence, and yet is caused to believe that *p* in some strange or bizarre way. And in these cases don't we want to say that that person isn't reasoning rationally?

Foley simply bites the bullet here and claims our intuitions are misguided. He defends this view by asking us to consider two hypotheticals. In each a person *S* has evidence *e* to believe that *p*, and recognizes *e* to be such evidence. In the first case *e* causes *S* to believe that *p* while in the second case something else does. But both cases are, so to speak, phenomenologically identical for *S*. Roughly, Foley's defence is this: since each hypothetical in some sense *feels* the same to *S*, *S* must be equally rational (or irrational) in each. Further, since Foley notices that such hypotheticals can be constructed for any causal requirement one might think of, he feels a quite strong conclusion follows. 'Namely', he says, 'there is *no* epistemic sense of rational belief in which a necessary condition of *S*'s rationally believing *p* is that his belief have an appropriate causal history ... (A)nd this is so regardless of any differences there might be in the histories of the beliefs in the two situations' (pp. 116 and 121).

In contrast, I think that by using his own method we can see that Foley's views on this matter are wrong. For example, consider the following two hypotheticals. In the first there is a person *S* going through a bit of reasoning consisting of three stages. In the first stage *S* finds herself thinking about a particular set of her beliefs $|\exists$. The second stage is 'directly' caused by the first and is her realization that $|\exists$ is adequate evidence to believe *p*. The third stage is similarly caused by the previous two and consists in *S* coming to believe that *p*. So the first hypothetical is an example of justified reasoning to the belief that *p*.

Now consider the second hypothetical. By hypothesis, *S* wouldn't be able to distinguish between the two were we able to transport her from one to the other. Indeed, we may even suppose that the two situations are at all times identical in terms of *S*'s brain-states. What does differ between them is the nomic facts involved. In the second hypothetical we find a rapid monkey, furiously hitting the keyboard of a computer so that each strike of a key causes a new brain-state in *S*. Further, *only* key-strikes cause *S*'s brain-states to change in this hypothetical. Of course *S* isn't aware that her beliefs and desires are being completely manipulated. For that matter, neither is the monkey. Yet Foley's view requires us not only to say that *S*'s beliefs are rational, but that *S is being* rational in both situations. Obviously though, no one is being rational in the second situation.

The point is that reasoning involves a certain sort of causal autonomy on the part of the cognizer. Otherwise, there is no reason to think of the person as reasoning at all. So, for example, though *S* may believe in both hypotheticals that she is inferring rationally, this belief will be true in the first and false in the second; and though it is not true that *S* is being irrational in the second situation, she's just not reasoning at all. So we can now see the problem with Foley's views. In fact *any* epistemic sense of rational belief has a

SOCIAL PHILOSOPHY AND POLICY

Edited by Ellen Frankel Paul

Each 192-page issue of **Social Philosophy and Policy** is devoted to a single topic in social or political philosophy, with specially invited papers (and, for designated issues, unsolicited papers) by leading scholars on the most important issues of our time. The topics for Volume 4 are **Philosophy and Law,** and **Philanthropy.** Previous issues have focused on **Distributive Justice, Human Rights, Liberty and Equality, Ethics and Economics, Nuclear Rights/Nuclear Wrongs,** and **Marxism and Liberalism,** with contributions from such distinguished philosophers as Maurice Cranston, Ronald Dworkin, R. M. Hare, David Gauthier, J. R. Lucas, Amartya Sen, Jon Elster, and G. A. Cohen, among others.

Philosophy and Law contains the following articles:

Mercy and Justice
JEFFRIE G. MURPHY

How promising works and how some laws create moral obligations: Analogous mechanisms from a utilitarian point of view
DONALD H. REGAN

Reflections on originalism
DAVID LYONS

Choosing a legal theory on moral grounds
PHILIP SOPER

Wrongful life and the counterfactual element in harming
JOEL FEINBERG

Taxation in a lockean world
RICHARD EPSTEIN

Justice in settlements
JULES COLEMAN and CHARLES SILVER

Practical wisdom and professional character
ANTHONY KRONMAN

October and April for the Social Philosophy and Policy Center, Bowling Green State University, USA

Subscription to Volume 4, 1986/7
Individuals: £11 (UK) £12.50 (overseas) US$15 (N. America & Japan)
Institutions: £26 (UK) £31.50 (overseas) US$45 (N. America & Japan)

Please send orders with payment to:

Iris Taylor
Journals Department
Basil Blackwell
108 Cowley Road
Oxford OX4 1JF
England

or to: Journals Department
Basil Blackwell
PO Box 1320
Murray Hill Station
NY 10156
USA

Payment can be made by cheque or money order made payable to Basil Blackwell Ltd, by credit card (Access, American Express, Barclaycard, Diners Club, Mastercharge, Visa; please state card number and expiry date), or through the National Girobank (account number 236 6053). Sample copies, single issues are available; please write for details.

Basil Blackwell · Oxford and New York

The Philosophical Review

A Quarterly Journal Edited by the Sage School of Philosophy

July 1986

The Passivity of Emotions .... *Robert M. Gordon*

The Unity of Descartes's Man .... *Paul Hoffman*

Morality and Freedom: Kant's Reciprocity Thesis .... *Henry E. Allison*

220 Goldwin Smith
Cornell University
Ithaca, New York 14853-3201

U.S. Subscription Rates:	*Rates outside the U.S.:*
$12.00 per year for students, the retired, and the unemployed	$17.00 per year for students, the retired, and the unemployed
$20.00 per year for other individuals; $6.00 per single issue	$25.00 per year for other individuals; $8.00 per single issue
$34.00 per year for institutions; $ 9.00 per single issue	$39.00 per year for institutions; $11.00 per single issue

VISA and MasterCard Accepted

INQUIRY *Editor: Alastair Hannay*

Vol. 29 (1986), No. 2
Symposium on Patricia Smith Churchland's *Neurophilosophy*
Editor's Introduction
I. Can Intuitive Psychology Survive the Growth of Neuroscience? — Keith Campbell
II. Neurophilosophy: A Principled Sceptic's Response — Geoffrey Madell
III. Nemo Psychologus nisi Physiologus — K.V. Wilkes
IV. Explaining Behavior: Bringing the Brain Back in — Christine A. Skarda
V. Brain-mind Philosophy — Aaron Smith
VI. A Program for the Neurobiology of Mind — Martin Sereno
VII. Replies to Comments — Patricia Smith Churchland
Following and Not Following Wittgenstein (J.N. Findlay, *Wittgenstein: A Critique*; Saul A. Kripke, *Wittgenstein on Rules and Private Language*; Colin McGinn, *Wittgenstein on Meaning*) — John Llewelyn
The Importance of Ownership (Alan Ryan, *Property and Political Theory*) — Onora O'Neill

Subscription price 1986:
Institutions: USD 53.00 Individuals: USD 27.00

Published quarterly by Norwegian University Press, P.O. Box 2959 Tøyen, 0608 Oslo 6, Norway — Publications Expediting Inc., 200 Meacham Ave., Elmont, NY 11003, USA. Editorial address: Institute of Philosophy, P.O. Box 1024 Blindern, 0315 Oslo 3, Norway.

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, T. R. Baldwin; T. S. Champlin, M. Clark, J. P. Dancy, Andrew Harrison, B. J. Heal, R. F. Holland, D. H. Mellor, H. W. Noonan, P. Smith, M. A. Stewart, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. ANALYSIS is published by Basil Blackwell Ltd, 108 Cowley Road, Oxford OX4 1JF, England and P.O. Box 1320, Murray Hill Station, NY 10156, U.S.A. Each volume comprises four numbers, each of at least 48 pages, appearing in January, March, June and October. Orders should be sent to the publisher. Subscriptions prices for 1986 are as follows:

	U.K.	Overseas (incl. Eire)	U.S.A. & Canada
Institutions	£12.00	£15.00	U.S. $33.00
Individuals	£8.00	£10.00	U.S. $21.00

BACK ISSUES. Volumes 1–37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

MICROFILM copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30–32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

ADVERTISING. For details contact Bill Drake, telephone (0869) 38477, or write c/o the publishers.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributions are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain

Vol. 47 No. 2 (New Series No. 214) March 1987

ANALYSIS

Edited by

Christopher Kirwan

CONTENTS

Editorial	
On a persistent fallacy regarding the *de re*	Thomas E. Patton
Davidson on force and convention	William K. Blackburn
Appiah on 'if'	V. H. Dudman
Deontic trouble in speech act botany	Savas L. Tsohatzidis
Prior's theory of truth	Charles Sayward
Quantification, description, and identity	Herbert Hochberg
Morris on identity	George Englebretsen
The puzzle of the flash stockman: a reply to David Lewis	Peter Lamarque
Thoughts and definitions	Justin Broackes
How many minds has a tredecim?	George Rea
Knowledge, belief and existence	C. J. F. Williams
Evidential insatiability	N. M. L. Nathan
Solipsism regained	Mark Leon
Two 'by' ways	Michael Wreen
Antiphon	Gerald Vision

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003-2638

BLACKWELL

Essays Metaphysical and Moral

Selected Philosophical Papers

J. J. C. SMART

This volume collects J. J. C. Smart's essays on a variety of philosophical topics. It includes a spirited defence of scientific realism and of a physicalist philosophy of mind, and is marked throughout by a thorough-going concern for plausibility and an extraordinary conceptual clarity.
320 pages, **£25.00** (0 631 15246 6)

Love and Human Separateness

İLHAM DILMAN

İlham Dilman undertakes a penetrating examination of the mind-body problem with particular emphasis on the implications that our view of this has for understanding what it means to know other people. He approaches his main theme by considering the Cartesian divorce between mind and body, as well as discussing what literature has to say on the question of sexual love.
176 pages, **£17.50** (0 631 15313 6)

Language and Reality

An Introduction to the Philosophy of Language

MICHAEL DEVITT and KIM STERELNY

This book is an introduction to the complex relations between language and the world and between language and mind. The book surveys and takes issue with a number of recent developments in philosophy of language and provides detailed coverage of a difficult and vital area of philosophy.
288 pages, hardback **£27.50** (0 631 15011 0)
paperback **£8.95** (0 631 15012 9)

Convention

DAVID LEWIS

New in paperback

Convention was immediately recognized as a major contribution to the subject and its significance has remained undiminished since its first publication in 1969. It is now once more available, and published in paperback for the first time.
'readers will be indebted to the author of this book' *Philosophical Quarterly*
224 pages, paperback **£8.95** (0 631 15079 X)

Counterfactuals

DAVID LEWIS

New in paperback

'This is an excellent book. It combines shrewd philosophical sense with a fine technical expertise. The statement of views is concise and forthright,' *Mind*
This classic text is now reissued.
168 pages, paperback **£7.00** (0 631 12497 7)

Basil Blackwell

108 Cowley Road, Oxford OX4 1JF
Suite 1503, 432 Park Avenue South, New York NY 10016

THE PHILOSOPHICAL QUARTERLY

Executive Editor: John Hall

Founded by the Scots Philosophical Club and the University of St Andrews in 1950, **The Philosophical Quarterly** aims to foster and publish significant contributions in every branch of the subject, promoting discussion of recent philosophical works, and providing critical reviews of recent books, in a style accessible to as wide a group of philosophers as possible.

Published January, April, July, October for the Scots Philosophical Club and the University of St Andrews
Subscription Rates Volume 36, 1986
Individuals: £11 (UK) £13.75 (overseas) US$26 (N America)
Institutions: £27.50 (UK) £34.35 (overseas) US$57.50 (N America)

Please send orders with payment to:

Iris Taylor
Journals Department
Basil Blackwell
108 Cowley Road
Oxford OX4 1JF
England

or to: Journals Department
Basil Blackwell
PO Box 1320
Murray Hill Station
NY 10156
USA

Basil Blackwell · Oxford and New York

SOCIAL PHILOSOPHY AND POLICY

Edited by Ellen Frankel Paul

Each 192-page issue of **Social Philosophy and Policy** is devoted to a single topic in social or political philosophy, with specially invited papers (and, for designated issues, unsolicited papers) by leading scholars on the most important issues of our time. The topics for Volume 4 are **Philosophy and Law,** and **Philanthropy.** Previous issues have focused on **Distributive Justice, Human Rights, Liberty and Equality, Ethics and Economics, Nuclear Rights/Nuclear Wrongs,** and **Marxism and Liberalism,** with contributions from such distinguished philosophers as Maurice Cranston, Ronald Dworkin, R. M. Hare, David Gauthier, J. R. Lucas, Amartya Sen, Jon Elster, and G. A. Cohen, among others.

October and April for the Social Philosophy and Policy Center, Bowling Green State University, USA
Subscription to Volume 4, 1986/7
Individuals: £11 (UK) £12.50 (overseas) US$15 (N. America & Japan)
Institutions: £26 (UK) £31.50 (overseas) US$45 (N. America & Japan)

Please send orders with payment to:

Iris Taylor
Journals Department
Basil Blackwell
108 Cowley Road
Oxford OX4 1JF
England

or to: **Journals Department**
Basil Blackwell
PO Box 1320
Murray Hill Station
NY 10156
USA

Basil Blackwell · Oxford and New York

ANALYSIS 47.2 MARCH 1987

EDITORIAL

I am pleased to announce on behalf of the Analysis Committee that the Committee have appointed Mr Peter Smith to succeed me as Editor after the current volume is complete in October 1987. I shall hand over to Mr Smith during July and August; in order to smooth the transition, it will be helpful if contributions submitted after 31 May are sent to:

Peter Smith,
Department of Philosophy
Hugh Owen Building,
University College of Wales,
Aberystwyth,
Dyfed SY23 3DY,
Wales.

ON A PERSISTENT FALLACY REGARDING THE *DE RE*

By THOMAS E. PATTON

I INTRODUCTION

MOSTLY as a legacy of W. V. Quine's jaundiced view of it in 'Quantifiers and Propositional Attitudes', I suspect, and later in *Word and Object*, the *de re*, transparent or relational idiom for indirect quotation and the ascription of propositional attitudes still often arouses mixed feelings of mistrust and disrespect. Such feelings are a surprise in David Kaplan, however, whose 'Quantifying In' analyses that idiom primarily in terms of a *de dicto*, opaque or notional idiom with which he is happy. While perhaps making it derivative, such an analysis must bestow upon the *de re* idiom whatever respectability attaches to the language which renders it. Yet in 'Demonstratives' Kaplan is heard voicing the familiar doubts. This could be missed, for he aims his doubts at what he calls a *pseudo* form of that idiom. But this shift to a new target can hardly be upheld, for Kaplan's examples of the pseudo form are in fact perfectly central cases of the genuine, the one and only, *de re*.

I need not insist on this last. But even a pseudo *de re* idiom, if such a distinct species there be, should not be impugned for a bad reason. Some years ago I defended some *de re* idiom against a Quinean argument to a shocking oddity about it, claiming to find a major fallacy (see [5]). Now, an experience of *dejà vu* while rereading an argument whereby Kaplan supports his doubts leads me to mount a similar defence. For I find a major fallacy in this argument too, and at least broadly (though I will not try to show this here), Kaplan's mistake and Quine's are the same. Exposing this fallacy should, I think, go a long way toward dispelling doubts about the *de re*. And a few comments on Kaplan's surrounding discussion may further advance that noble crusade.

II KAPLAN'S ARGUMENT

The relevant passage in Kaplan is a footnote, which I quote in full:

> 37.1 There is another form of common speech which may be thought to suggest formalization by quantifying in. I call this form the *pseudo de re*. A typical example is, 'John says that the lying S.O.B. who took my car is honest'. It is clear that John does not say, 'The lying S.O.B. who took your car is honest'. Does John say $\ulcorner$d is honest$\urcorner$ for some directly referential term d which the reporter believes to refer to the lying S.O.B. who took his car? Not necessarily. John may say something as simple as, 'The man I sent to you yesterday is honest'. The reporter has simply substituted his description for John's. What justifies this shocking falsification of John's speech? Nothing! But we do it, and often recognize – or don't care – when it is being done. The form lends itself to strikingly distorted reports. As Church has shown (*Introduction to Mathematical Logic*, p. 25), when John says 'Sir Walter Scott is the author of *Waverley*' use of the *pseudo de re* form (plus a quite plausible synonymy transformation) allows the report, 'John says that there are 29 counties in Utah'. I do not see that the existence of the *pseudo de re* form of report poses any issues of sufficient theoretical interest to make it worth pursuing. ([3] p. 116)

Kaplan's penultimate sentence introduces the argument in which I find a fallacy. Despite appearances, this is not Church's argument, for his result concerns no form of indirect quotation or propositional attitude ascription. Rather, Kaplan must be read as building a certain argument of his own, about the first of these topics but, as we will see, easily adapted to the second, around what Church has shown. Here is the relevant passage in Church:

> Again the sentence 'Sir Walter Scott is the author of *Waverley*' must have the same denotation as the sentence 'Sir Walter Scott is the man who wrote twenty-nine Waverley Novels altogether', since the name 'the author of *Waverley*' is replaced by another name of the same person: the latter sentence, it is plausible to suppose, if it is not synonymous with 'The number, such that Sir Walter Scott is the man who wrote that many Waverley Novels altogether, is twenty-nine', is at least so nearly so as to ensure its having the same denotation; and from this last sentence, in turn, replacing the complete subject by another name of the same number, we

> obtain, as still having the same denotation, the sentence 'The number of counties in Utah is twenty-nine'. ([2] p. 25)

And here, with obviously harmless simplifications, are the premises, intermediate lines and conclusion of the argument Kaplan must have in mind:

(1) John said that Scott wrote 29 books.
(2) John said that the number of books Scott wrote is 29.
(3) The number of books Scott wrote = the number of counties in Utah = 29.
(4) John said that the number of counties in Utah is 29.
(5) John said that there are 29 counties in Utah.

His premises are (1) and (3), while (2) and (4) are intermediate lines and (5) is his conclusion. Two rules of inference are at work here. One is Leibniz's Law, which gets us from (2) to (4), given (3). The other rule I will call Synonymy. It gets us from (1) to (2), and also from (4) to (5). Where *A* is an indirect quotation or propositional attitude ascription and the *complement* of *A* is the clause following the word 'that' in *A*, we can state Synonymy thus: if *A* is true and *A'* is like *A* except for having a synonymous clause in place of the complement of *A*, then *A'* is also true.

We should be clear at the outset how Kaplan views this argument. As his references to shocking falsification and striking distortion make plain, he is not presenting it as a valid one. His whole point is that its conclusion could be false when its premises were true. Rather, he is casting this argument as an *absurdum* in a *reductio* against his pseudo *de re* idiom. As a twist on this familiar methodology, however, it is not a premise that gets reduced to absurdity, but rather a rule of inference. For the essence of his *de re* idiom, or at least its key feature here, is a complete openness of its singular terms to Leibniz's Law. Since the only other rule Kaplan's argument employs is 'a quite plausible synonymy transformation', only that Law can be blamed for our being able to reach invalid conclusions like (5). And we can now fault the idiom itself for giving Leibniz such free rein.

By contrast, while agreeing with Kaplan that his argument is invalid, I will put the blame for our being able to reach (5) on Synonymy rather than on Leibniz's Law. For such rules need rationales, and I will argue that in our context, Synonymy can merely appear to have one. Calling the lines to which these rules apply in arguments like Kaplan's their *sources*, the appearance owes to the fact that in Synonymy sources in his argument, grammatical form, as Russell would say, is misleading as to logical form. And its mereness, or the nonreality, owes to an unmeetable formal demand of employing both of these rules in such arguments, namely, that the sources and the results of both Synonymy and Leibniz's Law must all exhibit a single generic logical form. As a slogan for Kaplan's fallacy, I will charge him with changing logical form in midstream. Let us now see what basis can be developed for that charge.

III TWO RULES WITH CLASHING RATIONALES

Such claims as that (2) comes from (1) by Synonymy, say, or that (4) comes from (2) by Leibniz's Law, must find their ultimate support, I take it, in a familiar move to symbolic sentences and, employing their symbolism, formal rules of inference. We could hardly understand such claims if their makers were to insist that no such support need be in prospect. I take it too that rationales for Kaplan's two rules must be ultimately couchable in terms of logical form, conceived as that which symbolization bares. This would hold for such rules as Universal Instantiation, say, in quantificational logic. But seeking rationales for Kaplan's two rules, we will find that each calls for a particular logical form in its sources, while implying that form for its results, where these forms differ, for the two rules, so that no one sentence can exhibit both. Yet (2), in Kaplan's argument, must be both a result of Synonymy and a source for Leibniz's Law, while the reverse holds for (4). These are dual roles which (2) and (4) can merely appear to play.

Quine rationalizes Leibniz's Law thus: 'When a singular term is used in a sentence purely to specify its object, and the sentence is true of the object, then certainly the sentence will stay true when any other singular term is substituted that designates the same object' ([7] p. 142). Earlier, he speaks of a term's being 'used as a means simply of specifying its object ... for the rest of the sentence to say something about ...' ([7] p. 141). This invokes a model of what it at least can be to make a statement: the speaker *refers* to one or more objects, using singular terms for this, and, using 'the rest of the sentence', identifiable as a *predicate*, he says something about them. The truth-value of his statement will depend on what these objects are, and on what he says about them, or on his predicate, but not on how he refers to them. Leibniz's Law will obviously hold for singular terms used as per this model. But when it holds for a singular term *T* in a sentence *S*, yielding a sentence *S'* which is like *S* except for having tokens of a term *T'* in place of one or more tokens of *T*, its holding in that way presupposes, as per its rationale, a particular *subject-predicate* (SP) *analysis* of *S*, one with *T* as subject and 'the rest of' *S* – or better, a logician's predicate gotten from *S* by putting tokens of a new variable for the tokens of *T* that get replaced in forming *S'* – as predicate. For example, when (4) comes from (2) by Leibniz's Law, this SP analysis of (2) casts 'the number of books Scott wrote' as subject and 'John said that *x* is 29' as predicate. (Similarly, a step of Existential Generalization from '*Gab*' to '$\exists xGxb$', say, presupposes, as per our rationale for that rule, an SP analysis of '*Gab*' on which '*a*' is subject and '*Gxb*' is predicate. Quantificational semantics in whatever form, seeking a truth-value for '$\exists xGxb$', will deal somehow with this predicate.)

Now SP analyses correspond many-one to logical forms. (Thus '*Gab*', which has a single logical form, has three SP analyses: the one just given, one with '*b*' as subject and '*Gay*' as predicate, and one with '*a*' and '*b*', taken in that order, as subjects and '*Gxy*' as predicate.) I will say that a logical form as fixed by a symbolic sentence *S allows* each SP analysis of *S*. Then by the transitivity of presupposition, Leibniz's Law, in holding for *T* in *S* as above, presupposes not only an SP analysis Q of *S* which casts *T* as subject, but also a logical form for *S* which allows Q.

And as a rationale for Synonymy, can we not simply say that this rule reflects the very spirit of indirect quotation? A little reflection should convince us, I think, that no other rationale for this rule is in sight. In directly quoting John, we must utter his actual words, but in indirectly quoting him, we need only utter words which, in a suitable sense, mean the same as his own did. This sense of meaning the same must cover more than is usually called synonymy, if that notion relates to *descriptive* rather than to *demonstrative* conventions, as J. L. Austin distinguishes these two in 'Truth'. (Here we enter the conceptual seas which Kaplan's 'Demonstratives' charts so ably.) And Synonymy allows itself some of this further coverage too, given that proper names like 'Scott' and 'Utah' have tacit demonstrative components. Now for my purposes here we can safely oversimplify, stretching a word, and just call it synonymy which indirect quotation and Synonymy both need. But a certain difference as to how they need it will prove crucial not to ignore.

For this and upcoming points a little jargon will be helpful. I will henceforth call indirectly quoting, viewable as a specialized speech-act type, Indirect Quotation. While dropping the capitals, indirect quotations will just be sentences like (1), say, in Kaplan's argument. As with Russell's denoting phrases in 'On Denoting', indirect quotations are a purely syntactic category, and, as denoting phrases can coherently fail to denote, could coherently fail to be reachable by Indirect Quotation even when true.

In these terms, the crucial difference is that Indirect Quotation needs synonymy between the complement of an indirect quotation and John's actual words, while Synonymy needs synonymy rather between two such complements. This difference means that if Synonymy, as a formal rule, is to reflect the spirit of Indirect Quotation, thus finding its rationale, then Synonymy must apply to indirect quotations which are reachable by Indirect Quotation. With no other rationale for it in sight, Synonymy can thus have a rationale only if its sources are indirect quotations of a style whose complements are guaranteed synonymy with John's actual words. But indirect quotations with complements whose singular terms are open to Leibniz's Law lack this feature. Yet in an argument cast as an *absurdum* in a *reductio* against a *de re* idiom, the indirect quotations will of course display this openness. This is what turns an otherwise symmetrical situation into one slanted against Synonymy. We can say first that no style of indirect quotation can permit rationales for both Synonymy and Leibniz's Law. We can then add that in our context, only the style which permits a rationale for that Law could be at hand.

As implied by my reference to an otherwise symmetrical situation, there is in fact a style of indirect quotation which permits Synonymy a rationale, namely, and not surprisingly, the *de dicto* style. And again, it is rather the *de re* style which permits a rationale for Leibniz's Law as arguments like Kaplan's employ it. Now we could sharpen the claims just made, which imply a clash of rationales, by symbolizing these two styles, thus baring logical form. Toward this, we can build on work already done for belief-ascriptions, either Quine's in 'Quantifiers and Propositional Attitudes' or Kaplan's in 'Quantifying In'. Following Kaplan, say, and adapting to indirect quotations by letting his 'B' and 'Bel' for the *de dicto* and *de re* idioms become 'S' and 'Say', respectively, one disambiguation of (4), in particular, yields:

(4a) John S 'the number of counties in Utah is 29'
(4b) John Say ('x is 29', the number of counties in Utah)

Similarly, we know what (1a), (2a) and (5a) would be, and can have:

(2b) John Say ('x is 29', the number of books Scott wrote)

Doing our best for Kaplan's argument, we can get from (1a) to (2a) by Synonymy, from (2b) and (3) to (4b) by Leibniz's Law, and from (4a) to (5a) by Synonymy. A *non sequitur* occurs between (2) and (4) and another between (4) and (5), a fallacy of equivocation for each of his tacit midstream changes of logical form. Granting that our 'S' and 'Say' versions do bare such form, his problem can be put thus. Synonymy and Leibniz's Law are both *replacement* rules. More narrowly, each replaces one singular term by another. But in general, and for Kaplan's (2) in particular, *de re* logical forms will not contain singular terms, namely, designations of sentences, of the kind that Synonymy replaces. And in general, and for the leftmost singular term in (3), in particular, which Kaplan uses Leibniz's Law to replace, *de dicto* logical forms will not contain singular terms found in the complements of indirect quotations. Thus neither of his two rules can take a result of the other as a source. Yet the only justifications in sight for the five lines of his argument will have this happening twice.

IV RESPONDING TO A FINAL DOUBT

If a *de re* idiom is accepted even tentatively, even for the unfriendly purpose of a *reductio* revealing more or less paradoxical consequences of its adoption, if such an idiom is even thus given enough rope to hang itself with, it must also bé given a logical form which makes intelligible the openness to Leibniz's Law of singular terms in its complements. While Kaplan and Quine both recognize this in their symbolisms for *de dicto* and *de re* senses, their intuitions as to logical relations, among the indirect quotations and belief-ascriptions which those symbolisms let us disambiguate, seem at crucial points to ignore it. The ambiguous English 'that'-clause is surely the villain here. Had Kaplan rehearsed his argument giving either an 'S' form or a 'Say' form to every line but (3), would he still have let Synonymy, 'plausible synonymy transformation' though it may be, mix freely with Leibniz's Law in reaching his conclusion? (We could similarly ask about Quine's fallacious argument, whose counterpart to Synonymy is a rule we might call Logical Acumen.)

With logical form so crucial, however, a doubt might arise touching my appeal to it here. For what kind of symbolism should we have in mind in identifying logical form as that which symbolization bares? If this is the symbolism of elementary logic texts, then (4a) and (4b), say, do not bare logical form. At best we can claim that they *fix* it, assuming fixed ways of treating definite descriptions and quoted expressions in that symbolism. But it is easy if tedious to see that Kaplan's *non sequiturs* survive a descent to the elementary level. Such a descent even shows that his argument, so far from merely employing a flawed strategy, is a doomed enterprise. However, my reasons for these findings of survival and doom must be omitted here.

University of British Columbia,
1866 Main Mall, Vancouver,
Canada V6T 1W5

© THOMAS E. PATTON 1987

REFERENCES

[1] J. L. Austin, *Philosophical Papers* (London: Oxford University Press, 1961).
[2] Alonzo Church, *Introduction to Mathematical Logic* (Princeton: Princeton University Press, 1956).
[3] David Kaplan, 'Demonstratives', mimeographed 1977 and circulated.
[4] David Kaplan, 'Quantifying In', in *Words and Objections*, Donald Davidson and Jaakko Hintikka, editors (Dordrecht: D. Reidel, 1969).
[5] T. E. Patton, 'Is Transparent Belief "Intolerably Odd"?', *Dialogue*, XIII (1974).
[6] W. V. Quine, 'Quantifiers and Propositional Attitudes', *The Journal of Philosophy*, 53 (1956).
[7] W. V. Quine, *Word and Oject* (Cambridge: MIT Press, 1960).
[8] Bertrand Russell, *Logic and Knowledge*, Robert C. Marsh, editor (London: Allen and Unwin, 1958).

DAVIDSON ON FORCE AND CONVENTION

By William K. Blackburn

THE idea is familiar that there is some relation, perhaps a conventional one, between illocutionary force and the various English moods. Thus declarative sentences are related to acts of assertion, interrogatives to questions, and so on. But it is not clear exactly what the relation here is. One natural suggestion (hereafter called 'the thesis') is that it is part of the *conventional meaning* of, e.g. a declarative sentence, that it is used in 'normal' or 'standard' circumstances as an assertion.

In 'Communication and Convention', Donald Davidson argues against the thesis (D. Davidson, *Inquiries into Truth and Interpretation*, Oxford University Press 1984, pp. 265–80). One of his arguments is particularly interesting because, if sound, it shows that *any* attempt to explain the meaning of sentences in terms of the conventions that govern their use in standard conditions must fail. Davidson writes:

> the ... thesis [that there is a convention that under 'standard' conditions the utterance of an imperative *is* an order] can't explain the difference in meaning between an imperative and a declarative sentence, a difference which exists quite independently of illocutionary force. The ... thesis can't explain this because it postulates a convention that is in force only under 'standard' conditions. You can't *use* a convention by breaking it, you can only *abuse* it. But the difference between an imperative and a declarative can, and very often is, quite properly used in situations where mood and illocutionary force are not 'standard' ....
>
> The gist of this remark applies also to the kind of theory that tries to derive the literal meaning of each sentence from a 'standard' use. Since the literal meaning operates as well when the use is absent as when it is present, no convention that operates only in 'standard' situations can give the literal meaning. (p. 275)

As I interpret it, the argument in this passage can be summarized as follows:

(1) A sentence (e.g. a declarative) has its full literal meaning even when it is not used in a 'standard' way (i.e. as an assertion).

(2) Therefore, no convention that operates only in standard situations gives (any part of) the meaning of the sentence.

(3) Therefore, it is not part of the conventional meaning of, e.g., a declarative sentence, that it is used in standard situations to make an assertion. The thesis is false.

Let us grant that (1) is correct. I shall argue that, depending on how we understand (2), either (2) does not follow from (1) or (3) does not follow from (2).

To see this, we must distinguish two kinds of situation in which a convention fails to operate. Consider an analogy. A *law* may not operate for at least two quite different sorts of reason. On the one hand, a law may simply not exist, or no longer exist. Thus it was once against the law to purchase liquor in many parts of Canada and the United States, but these laws have now been repeated. Similarly, when a court throws out a law that it judges to be unconstitutional, that law is no longer in force, i.e. it ceases to exist.

By contrast, it may be *part of the content* of a law that it applies only in certain situations. For example, it is forbidden to turn left at certain intersections between 4 and 6 p.m. This law is still *in force* at 7 o'clock, but it does not *apply*.

There is a similar distinction between two kinds of situation in which a convention does not operate. On the one hand, the convention may not exist, or no longer exist. An example is the obsolete convention among English speakers of using the word 'thou' as the second-person singular personal pronoun. To say that this convention no longer exists is to say that people no longer follow it; their behaviour is no longer governed by it and cannot be explained by citing it.

On the other hand a convention may continue to exist, but not apply in a particular context. Imagine a community with the following convention: they greet each other by saying 'Hello' on weekdays, but they use some other expression as a greeting on Sundays. In this case, the convention of using 'Hello' as a weekday greeting *exists* on Sundays, but it does not *apply*. Notice that a person's behaviour can be explained by the convention even when it does not apply. A woman might say 'Hello' because she mistakenly believes it is a weekday; in that case, her behaviour is explained in part by citing the convention. Thus even on Sundays it is a convention, in this community, to use 'Hello' as a weekday greeting. But it is *part of the content* of the convention that it does not apply on Sundays.

Let's return to Davidson's argument. We can grant that a declarative sentence retains its full meaning when it is not used to make an assertion. More generally, sentences keep their literal meaning when they are not used in a 'standard' way, so premise (1) is correct. Does it follow that (2) no convention that operates only in standard situations gives the meaning of the sentence? That depends on what we mean by 'operate'.

If we mean that no convention that *exists* only in standard situations gives the meaning of a declarative sentence, (2) obviously does follow from (1). Meaning-conventions do not cease to exist when conditions are non-standard. But, on this interpretation, (3) does not follow from (2). Of course, if we understood the thesis to be that the meaning of a declarative sentence is determined by conventions that exist only in standard situations, Davidson's argument would be persuasive. But it is more natural, and more charitable, to interpret the thesis as saying that it is *part of the content* of the conventions that give the meaning of declarative sentences that, in normal circumstances, these sentences are used to make assertions. So, on the only reasonable construal of (3), it does not follow from (2). The conventions that are supposed to give the meaning of declarative sentences are *not* conventions that exist only in standard situations.

On the other hand, if by (2) we mean that no convention which *applies* only in standard situations gives the literal meaning of the sentence, this does *not* follow from the fact that the sentence retains its literal meaning when it is used in a non-standard way. If we take seriously the view that the meaning of a sentence is given by the conventions that govern its use, there is no reason to exclude those conventions that specify as part of their content that they apply only in certain circumstances.

I conclude that, on either interpretation, the argument is unsuccessful. It fails to show that the meaning of a sentence is not given by the conventions that govern its use in standard situations.

St Mary's University,
Halifax, Nova Scotia,
Canada B3H 3C3

© WILLIAM K. BLACKBURN 1987

APPIAH ON 'IF'

By V. H. DUDMAN

SUPPOSE a plurality of urns containing little coloured balls, among them urn *B* containing 999,999 white balls and one black one. And suppose that, knowing this, we choose a ball *z* from an unknown urn, possibly *B*, with our eyes shut. Then, says Anthony Appiah (in Part III of his *Assertion and Conditionals*, Cambridge University Press 1985: see p. 172), we ought to assent when someone says

(1) If z is from urn *B*, *z* is not black.

Now I am very sorry, but the hard fact is that we do *not* assent to (1). Not in Australia, anyway, where I have rehearsed Appiah's scenario to literally dozens of my fellow-citizens, mostly students. All, all have responded, sometimes with passion, that anyone acquiescing in (1) would have to be ignoring the black ball. They concede that if *z* is from *B* it is 'almost certainly' not black, but exceptionlessly insist that (1) remains strictly untenable so long as there is a black ball in urn *B*, even if the number of accompanying white balls is increased billionfold. Appiah contends that one ought to assent to (1) 'because there is only the very faintest chance of your getting a black ball out of urn *B*' (p. 172); but my respondents counter that it is precisely because there is a chance, however faint, of drawing the black ball from *B* that one ought *not* assent to (1). Evidently only those who believe none of *B*'s balls black are entitled to assert (1).

I conclude that Appiah's proposals for (1) are mistaken and a key tenet of his book consequently false. The tenet, called Adams' Hypothesis after Ernest Adams, is that an 'if'-sentence like (1) is assertible exactly when the conditional probability of its consequent, given its antecedent, is high (e.g. p. 4, p. 174). Yet (1) remains unassertible even after Appiah (pp. 171-4) has contrived an extremely high conditional probability for it. It would seem that Appiah has obligingly devised a counterexample to his own theory.

(My own theory for 'if'-sentences like (1), with a constituent sentence introduced by 'if', is that 'if' then serves to announce the antecedent proposition, burden of that constituent sentence, as a *hypothesis*. Reasons for its being thus taken for granted vary considerably; but in (1)'s case, I contend, the reason for confessing to the hypothesis that *z* came from urn *B* is so as to be able to use it as a premiss for inferring the consequent proposition that *z* is not black. Seeing that this conclusion follows only given that there is no black ball in urn *B*, I take the earnests of my respondents, cited above, as confirming my account as well as overturning Appiah's.)

An 'if'-sentence is 'like (1)' for the purpose of Adams' Hypothesis when it comprises two 'ordinary indicative sentences', one introduced by 'if' (p. 162). The two (ordinary indicative) sentences are an absolute necessity here, because 'probability is probability of truth' (p. 174n), and it takes a sentence to express a candidate for truth. The 'probability of the antecedent' is spectral unless there is a sentence in the scope of 'if', then.

In defence of his claim that we ought to assent to (1), Appiah invokes a comparison with

(2) If I walk on my stairs I will not fall through them

which, he says (p. 172), one affirms notwithstanding the very remote possibility that his stairs have been secretly ravaged by termites. But the comparison is illicit, for (2) is not like (1) in our defined sense. Given the intended reading of (1), the string '*z* is from urn *B*' is understood exactly as if encountered standing alone between full stops: it is parsed as a sentence. But there is no sentence in the scope of 'if' when (2) occurs under its desired future interpretation. In particular, the string

(3) I walk on my stairs

never occurs as a sentence with future reference. The question whether what is affirmed about the future by (3) is probably true is forestalled by the problem that nothing whatever is affirmed about the future by (3). And therefore it is confused to invoke the probability (of the truth) of the antecedent in such a case as (2)'s. Adams' Hypothesis simply does not apply to 'if'-sentences like (2).

(Let it be perfectly clear that 'I *will* walk on my stairs' gets the 'antecedent' wrong for (2)'s case. *That* theory cannot cope with the undoubted difference between saying 'If it will be cooler after sunset I shall call back then' and saying 'If it is cooler after sunset I shall etc.'.)

My criticism so far can be condensed in two sentences. First, Appiah confuses (as 'indicative conditionals': see, e.g. p. 162 and pp. 170-1) 'if'-sentences like (2), to which Adams' Hypothesis does not apply, and 'if'-sentences like (1), to which it does. Secondly, when Adams' Hypothesis does apply it doesn't seem to work, for it gets the answer wrong in (1)'s case. I move now to some broader, less technical worries I have about Appiah's proposals for 'if'.

For involved reasons of a highly theoretical kind, Appiah concludes in Chapter 9 that 'if'-sentences like (1) 'do not have truth conditions' (e.g. p. 205). But it is easily shown that the sponsor of an 'if'-sentence like (1) says nothing true or false. If the burden of e.g. (1) were a candidate for truth then it would be possible both to deny it and to suppose it for the sake of argument, whereas in fact these facilities are both unavailable, in confirmation of which neither (4) nor (5) makes any sense:

(4) It is not the case that if *z* is from urn *B* *z* is not black
(5) Suppose that if *z* is from urn *B* *z* is not black.

Nor is it surprising that the burden of (1) should be neither true nor false. After all, sentences like (1) are grammatically congruent with sentences like

(6) Because/As/Since/Although/While *z* is from urn *B*, *z* is not black,

and the obvious reaction to (6) is that its speaker is up to something too *involved* to permit any such simple verdict as 'true'. Straining the terms, there is an antecedent and a consequent, both admittedly either true or false; but the overall burden of (6) is structured out of antecedent and consequent in such a way that its own truth or falsity is out of the question: there is too much going on.

With respect, while Appiah's reasoning in Chapter 9, which takes us deep into probability theory, might conceivably serve as a *proof* that 'if'-sentences like (1) have no truth conditions, it hardly provides an *explanation*. If the burdens of sentences like (1) are never of the true-or-false variety, that must surely be for some grander, simpler reason. – Such as the one just ventured for sentences like (6). And indeed careful listening discovers that the obstacle to (4)'s and (5)'s making sense is nothing but sheer unresolvable complexity in the burden of (1). It is not as if (4) and (5) flouted laws of probability.

I must not leave the impression that Appiah is without simpler inducements of his own for thinking that 'if'-sentences like (1) lack truth conditions. If 'if'-sentences like (1) have truth conditions, he reasons, 'there ought to be nothing problematic about their behaviour when they lie within the scope of truth-functional connectives' (p. 205). So he sets about uncovering problems (pp. 205–9), thereby advancing 'concrete semantic evidence' (p. 205). The program is surely coherent, but alas it commits its author to identifying truth-functional connectives in English; and here he meets with unsuccess. At one point, e.g., he is found invoking the assumption that

> 'or' is, as it seems to be, a truth function

(p. 208); and this assumption, even when confined to cases where 'or' separates two genuine sentences, is frankly heroic. Who undertakes to defend the claim that all four 'disjuncts' in (7) are candidates for truth?

(7) 'Then he can't be a pansy or he wouldn't cry over a girl.' 'He must be or he wouldn't cry at all.' [Muriel Spark 1960]

(A far more plausible theory, surely, is that 'or' in English, between sentences, subordinate clauses, nouns, adverbs, prepositions or whatever, demarcates *alternatives* – whether exclusive or not, and whether exhaustive or not.) The very idea that there might be English particles with truth-table meanings outruns this enquirer's credulity.

Another of my fundamental difficulties with Appiah's position is that I do not understand assertibility. It is evidently conceived (p. 5) as an observable trait of sentences, with Adams' Hypothesis explaining its possession for the particular case of 'if'-sentences like (1). But if asked whether, e.g.

(8) He has the merit – if it is one – of saying exactly what he means [E. M. Forster 1922]

was assertible, I should candidly not know what the issue was. I understand (8) itself well enough: its sponsor affirms someone's owning a certain merit while collaterally allowing that it might not really be a merit. Presumably, then, (8) will be asserted by speakers wanting to do exactly that. But what can be achieved by calling (8) assertible simpliciter? Having shed one strait-jacket, Appiah seems to me to have snuggled up in another: 'assertibility conditions' are no better than 'truth conditions' when it comes to coping with the essential complexity I descry in the burdens of 'if'-sentences like (1).

And another thing. Probabilistic proposals for 'if'-sentences like (1) impress me as downright implausible. Take (8) again. Primâ facie, neither the probability that speaking one's mind is a merit, nor the probability that 'he' has the merit of speaking his, has the remotest bearing upon what we have seen the sponsor of (8) to be up to. Again, assigning probabilities to pronouncements such as that we must not act save on a certainty, and that we ought not to act upon religion, can hardly be easy for most of us; yet it is easy for all of us to follow the reasoning in (9):

(9) If we must not act save on a certainty we ought not to act upon religion, for it is not certain [Blaise Pascal, d. 1662]

We need not necessarily agree with the reasoning, I suppose. But if we disagree, it will not be about probabilities.

Or take (10):

(10) If it really is a law of nature that people resemble the pets they keep, then what is waiting at home for the Archdeacon is a cute little toy poodle.

Here, plainly, we have an artful way of describing the Archdeacon. What is not plain is how probability could possibly have a hand in it.

Nor do I see how the probabilistic story might deal with cases where the antecedent is a deliberate falsehood, as when (11) is asserted by someone glancing into a patently empty room:

(11) If Her Majesty is in this room she's invisible.

What sort of bearing can the probability have that Her Majesty is in the room when it is understood on all sides that she is not?

Sentences like (1) are also used for avowals, not on the face of it easily exponible in probabilistic terms either:

(12) I may have said so, my dear. But if I did, the other was what I really expected [Henry Green 1952]

The 'if'-sentences in (8) to (12) are of course all like (1) in our defined grammatical sense, and accordingly Adams' Hypothesis is meant to predict 'the actual linguistic behaviour of speakers' (p. 5) who use them. But how it does this remains opaque to me. I feel, too, that the onus is on Appiah here.

The reader has doubtless observed, however, that in each of (8) to (12) the antecedent is acknowledged as a hypothesis. Indeed the parenthetical account I proffered above for (1) holds exactly for each of (9) to (11): in all four cases, I contend, we have a *condensed argument*. The author of (9) even supplies the necessary co-premiss or *ground*. The point about (10) is that unfurling the condensed argument provides an oblique way of affirming its shadowy ground, so that the Archdeacon is thereby described impressionistically. In (11) the observational ground that there is no sign of Her Majesty in the room allies with the hypothesis that Her Majesty is nevertheless in the room to yield the deductive conclusion that Her Majesty is invisible.

On the other hand, neither (8) nor (12) is exponible as a condensed argument. By confessing to the hypothesis that saying what one means is a merit, the sponsor of (8) is able to leave that question open while still enjoying use of the term. The trick here is that, by taking on the antecedent as a hypothesis, a speaker can cover himself in case that antecedent is false, or a nuisance to defend. And similarly for (12): by saying 'if', the speaker is able to waive the issue about what he said earlier, while insisting on what he then meant. In neither case does consequent stand to antecedent as conclusion to premiss.

Here, then, is one particular empirical distinction – I could authenticate many – among the burdens of 'if'-sentences like (1). It is a distinction that goes unmarked, obviously, by anything in the sentences themselves: we are looking at *unsignalled information*, conveyed, according to me, by the hearer's reading it in. But I am trying to make a point against Appiah: that Adams' Hypothesis surely provides too monolithic an account of 'if'-sentences like (1) to allow for the sort of diversity we have just observed.

Part III of Appiah's engaging book is a philosopher's essay developing and exploring the probabilistic approach to 'if'-sentences like (1) – and encouraging its critical assessment. The philosophic spirit exults in free speculation, and philosophical speculation about 'if' accordingly suffers as a genre from persistent under-exposure to the empirical evidence. Suffers, because the question how 'if' works in English remains an inexorably empirical one. The only rational way of attacking it is the scientific way, which means *starting* with the teeming evidence: going out and collecting lots of it, quite unselectively at first, and then open-mindedly sifting through it. And yet, gentle reader, have you collected even a hundred English 'if'-sentences? Without bias? Copied them out from their various sources, and pored over them without preconceptions? If you have not, may I suggest you try it before submitting to Appiah's well-reasoned enticements?[1]

Macquarie University,
North Ryde, N.S.W.,
Australia, 2113

© V. H. DUDMAN 1987

---

[1] I am most grateful to Anthony Appiah for generous and enlightening private correspondence. My own account of 'if'-sentences like (1) is sketched in 'Parsing "If"-sentences' (*Analysis* 44.4, October 1984, pp. 145-53; see p. 147). The idea of unsignalled information is aired in 'Conditional Interpretations of "If"-sentences' (*Australian Journal of Linguistics* 4, 1984, 143-204; see pp. 187-95) which deals at length with 'if'-sentences like (2). It is perhaps worth observing that the 'if'-sentence in (12) above is hospitable to 'then' even though not exponible as a condensed argument.

DEONTIC TROUBLE IN SPEECH ACT BOTANY

By SAVAS L. TSOHATZIDIS

J. R. SEARLE has the following problem about permission (J. R. Searle, 'A Taxonomy of Illocutionary Acts', *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge 1979, p. 22): although permissions seem to him to be illocutionary acts of the same kind as requests, orders, prohibitions, suggestions, entreaties, etc., his classification of illocutionary acts does not seem capable of reflecting this fact, since permissions fail to exemplify the defining characteristic of the class of illocutionary acts to which he has allocated the others. Requests, orders, prohibitions, suggestions, entreaties, etc. are, according to Searle, *directive* illocutionary acts, that is to say, essentially, attempts to get hearers do things (or refrain from doing things). But, Searle observes, permissions do not seem to be directive illocutionary acts at all, since they are evidently not attempts to get hearers do things (or refrain from doing things). Therefore, the classification, as it stands, cannot reflect the strong feeling that they belong to the same natural class with all the other acts that are counted as directive.

Searle's response to this problem is not to review his classification, but to define permission in terms of certain other uncontroversially directive acts, and thus to conclude that, initial appearances notwithstanding, his classification does match his pretheoretical convictions. The definition has, at present, two variants (which are ensured to be equivalent by virtue of Searle's thesis that acts of forbidding that p are identical to acts of ordering that not p; J. R. Searle and D. Vanderveken, *Foundations of Illocutionary Logic*, Cambridge 1985, p. 202). According to the first variant ('A Taxonomy of Illocutionary Acts', p. 22) permissions that something be done are nothing else but 'illocutionary negations' of orders that it not be done (in other words, sentences like 'I give you permission to go' are equivalent to sentences like 'I do not order you not to go'). According to the second variant (*Foundations of Illocutionary Logic*, p. 202), permissions that something be done are nothing else but 'illocutionary negations' of prohibitions that it be done (in other words, sentences like 'I give you permission to go' are equivalent to sentences like 'I do not forbid you to go'). No matter which one of the variants you adopt, Searle suggests, you have reduced permissions to well established species of directive illocutionary acts; you can therefore confidently regard the classification as reflecting your pretheoretical conviction that permissions are illocutionary acts of the same kind as requests, orders, prohibitions, suggestions, entreaties, etc.

I propose to show three things: first that, even if Searle's definitions were correct, it would not follow that they are of any help in resolving his initial problem; second, that they are not, in fact, correct; and third, that they are not equivalent to each other.

Searle's claim that his definitions of permission, assuming them to be correct, succeed in showing that this act is, after all, a directive illocutionary act is very confused indeed. This claim assumes that the operation of illocutionary negation on a speech act does not change the character of this act, and in particular that the negation of an act of ordering is itself an act of ordering, or that the negation of an act of forbidding is itself an act of forbidding. But, as Searle himself had pointed out elsewhere, 'illocutionary negations in general change the character of the illocutionary act' (J. R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge 1969, p. 32), and this is especially clear when the illocutionary acts negated are acts of ordering or acts of forbidding. Someone saying

(1) I do not order you not to go.

or

(2) I do not forbid you to go.

can hardly be described as performing an act of ordering or an act of forbidding. He can only be described, as Searle himself had indicated in his general discussion of illocutionary negation, as *refusing* to perform an act of ordering or an act of forbidding, or, alternatively, as *denying* that an act of ordering or an act of forbidding is being performed. Therefore, if (1) and (2) are, as Searle contends, correct translations of (3),

(3) I give you permission to go.

what Searle can claim he has achieved by devising them is not that he has shown that acts of permitting are a special sub-class of acts of ordering or of acts of forbidding, but that he has shown that acts of permitting are a special sub-class of acts of refusing or of acts of denying. And since, as Searle well knows, refusals and denials are *not* directive illocutionary acts, he must conclude that his proposed definitions, if they are correct, achieve precisely the contrary of what he claims they achieve: instead of showing that permissions are, after all, directive illocutionary acts, they show that they are definitely *not* directive illocutionary acts.

Let us now examine whether Searle's analysis of permission is correct in itself, since someone might well accept it, without caring for the problems it creates for Searle's classification. To show that it is not, in fact, correct, one might show either that there are sentences of the form 'I give you permission to *p*' which do not have the same meaning as corresponding sentences of the forms 'I do not order you not to *p*' and 'I do not forbid you to *p*', or that there are meaningful sentences of the form 'I give you permission to p' to which there correspond no meaningful sentences of the forms 'I do not order you not to p' or 'I do not forbid you to p' (or conversely). The latter of these things can be established by pointing out that, whereas (4) and (5) are meaningful, (6) is not even grammatical:

(4) I do not forbid you to go at all.
(5) I do not order you not to go at all.
*(6) I give you permission to go at all.

The former of these things can be established by pointing out that, although (7) can be (and, indeed, standardly is) taken to mean that its hearer will be held responsible for not satisfying the speaker's stated wish if he stays at his house *without* marrying his daughter, (8) and (9) cannot be taken to mean that their respective hearers will be held responsible for not satisfying the speaker's stated wish if they stay at his house *without* marrying his daughter:

(7) I give you permission either to leave this house or to marry my daughter.
(8) I do not forbid you either to leave this house or to marry my daughter.
(9) I do not order you either not to leave this house or not to marry my daughter.

Thus, Searle's analysis of permission is no more correct than his claim that, if it was correct, it would resolve his classification problem.

Let us finally examine whether Searle was correct in believing that his alternative ways of analysing permission are equivalent, since one might be inclined to accept his reason for regarding them as such (in other words, his claim that acts of forbidding that *p* and acts of ordering that no *p* are one and the same thing), while admitting that neither can offer an acceptable analysis of permission. To show that they are not, in fact, equivalent, one might show, again, either that there are sentences of the form 'I forbid you to *p*' which do not have the same meaning as corresponding sentences of the form 'I order you not to *p*', or that there are meaningful sentences of the form 'I forbid you to *p*' to which there correspond no meaningful sentences of the form 'I order you not to *p*' (or conversely). As regards the latter point, it suffices to observe that, whereas (10) is meaningful, (11) is not even grammatical:

(10) I forbid your future children to stay in this country.
*(11) I order your future children not to stay in this country.

As regards the former point, it suffices to observe that, although the speaker of (12) can be (and, indeed, standardly will be) taken to mean that he challenges his hearers to prove their bravery by not leaving, the speaker of (13) can hardly be taken to mean that he challenges his hearers to prove their bravery by not leaving:

(12) I order those who are brave not to leave.
(13) I forbid those who are brave to leave.

Therefore, Searle's claim that his alternative analyses of permission are equivalent is no more true than his claim that they are correct, or that they satisfactorily resolve his classification problem.

The general conclusion to be drawn is, it seems to me, this: if, as Searle and many others believe, a reasonable condition of adequacy on speech act theories is that they correctly analyse, and, in so doing, reflect the natural affinities between the concepts of ordering, prohibiting, and permitting, Searle's theory of speech acts clearly fails to satisfy this condition.

University of the Witwatersrand,
Johannesburg, South Africa

© SAVAS L. TSOHATZIDIS 1987

PRIOR'S THEORY OF TRUTH

By CHARLES SAYWARD

I TRUTH LOCUTIONS

CONSIDER these truth locutions:

1. 'it is true that' prefixed to sentences;
2. 'true proposition';
3. 'true belief', 'true assertion', 'true statement', etc.;
4. 'true sentence'.

Call a truth locution L in this set *basic* if every true sentence in which any other truth locution in this set is used has as a paraphrase a sentence in which no truth locution other than L is used.

Is there a basic truth locution? Corresponding to each locution is a view that identifies that locution (or, in the case of 3, one of the locutions) as basic. On each of the views 2-4 the basic truth locution is a predicate. On view 1 the basic truth locution is a unary sentential connective.

Views 2-4 differ as to the nature of the basic truth-value bearers. Views 3-4 hold that in a world devoid of sentient creatures nothing is true. View 2 holds that even in a world devoid of sentience there would still be things true. A proponent of view 1 rejects the idea that basic truth-value bearers *exist*, which is not to deny that it is true that snow is white or that it is true that grass is green, etc.

II PRIOR'S VIEW

We make statements like these:

I agree with everything he said;
Some things he believes about Alice are true;
No contradictory proposition is true.

Prior views any such generality as paraphrasable in a language having quantifiers which bind variables in sentence position:

For all *p*, if he said that *p*, then I agree that *p*;
For some *p*, that *p* is about Alice and he believes that *p* and it is true that *p*;
For all *p*, if it is contradictory that *p*, then it is not true that *p*.

As long as such a language lacks locution 4 ('true sentence') the sentential connective 'it is true that' is the basic truth locution. This is a way of putting Prior's view.

Why the proviso? What prevents 'true sentence' from being treated like locutions 2 and 3? Prior has an explanation (1971, p. 100):

> No sentence in any language considered by Tarski can discuss its own truth or falsehood, or the truth or falsehood of other sentences of its own language. But we have allowed it as possible – given favourable circumstances – for people to say, think, or fear things about the truth and falsehood of what they say, think, or fear; e.g. a man may – given favourable circumstances – say, think, or fear on some occasion that something that he says, thinks, or fears on that occasion is false, and this will itself count as one of the things that he says, thinks, or fears on that occasion. But this in itself constitutes at most a difference from, not a conflict with, what Tarski says. For we are not using 'true' and 'false' in his sense – Tarski and we are not discussing the same subject.

Evidently Prior's reasoning goes like this: The Epimenides is formulable in a language which contains the means of designating its own expressions, the usual logical constants and a truth locution applicable to its own sentences. It is possible to prove a contradictory sentence of the langauge from commonplace premises. But the Epimenides is not formulable in a self-referential language with a sentential connective as its basic truth locution. Thus locution 1 is not basic relative to locution 4.

Why think that the Epimenides is not formulable in a self-referential language containing 'it is true that' as its basic locution?

A fundamental fact about this connective is that it is null or redundant:

For every *p*, it is true that *p* if and only if *p*

Since 'it is true that' is null, it is also dispensable.

This might have led Prior to think the Epimenides is not formulable in a language in which 'it is true that' is basic. For if the basic truth locution is redundant then everything sayable in the language is sayable in a subset which does not contain any truth locution. But the Epimenides is not formulable in the subset. So it is not formulable in the language.

On the other hand, it certainly seems as if 'it is true that' is basic relative to 'true sentence'. For consider that

> 'Snow is white' is true

is paraphrased by

> For some one *p*, 'Snow is white' expresses the proposition (or, says) that *p* and it is true that *p*.

The basic thought is that a sentence is true (false, undecidable) just in case it expresses one true (false, undecidable) proposition. And this latter concept is handled in a Priorean way. Using this basic idea I can think of no sentence using 'true sentence' not paraphrased by one using the connective.

III THE EPIMENIDES

Let us look at a simple form of the Epimenides. Consider

> (a) (a) is not true.

We confirm empirically that

> (b) (a) = '(a) is not true'.

But by a commonplace semantical principle it holds that

> (c) '(a) is not true' is true if and only if (a) is not true;

and from (b) and (c) it follows that

> (d) (a) is true if and only if (a) is not true,

a contradiction.

Now the principle underlying (c) is that every instance of the schema

> (T) x is true iff y

is true, where x is a designation of a sentence of the language and y is that sentence.

Now this general idea can be expressed in terms of 'it is true that': every instance of

> For some one *p*, x says that *p* and it is true that *p* if, and only if, y

is a true biconditional, where x and y satisfy the same conditions. By the redundancy property this schema reduces to this biconditional:

(P) For some one p, x says that p, and p if, and only if, y.

But now the paradoxical sentence (a) has a Priorean paraphrase:

(e) For some one p, (e) says that p, and not-p

We observe empirically that

(f) (e) = 'For some one p, (e) says that p, and not-p'.

And, applying the schema (P), we get this biconditional:

(g) For some one p, 'For some one p, (e) says that p, and not-p', says that p, and p if, and only if, for some one p, (e) says that p, and not-p.

From (g) and (f) we get this biconditional

(h) For some one p, (e) says that p, and p if, and only if, for some one p, (e) says that p, and not-p,

which is not a contradiction. For it is true if (e) expresses no proposition or if (e) is ambiguous, expressing more than one proposition. In fact, (h) is logically equivalent to this:

(i) It is not the case that, for some one p, (e) says that p

which is more than just possible.

IV CONCLUDING REMARKS

Was Prior right, then, in claiming that the sentential connective 'it is true that' is not basic relative to 'true sentence'? I do not think so.

Certainly the argument he gives fails to show this. It is true (as we have seen) that no contradiction follows from (f) and (g) and that a contradiction does follow from (b) and (c). Prior infers from this that a language containing the means of referring to its own expressions cannot contain its own truth predicate, while it can contain its own truth connective.

But this does not follow. The fact that (b) and (c) entail the contradiction means that they are not both true. Since (b) is empirically confirmed, (c) must be false. This in turn implies that the principle underlying (c) is false. That principle says schema (T)

x is true iff y

holds where x is a designation of sentence y. But it does not hold for sentence designation '(a)' and sentence (a).

This calls for a simple, obvious emendation. The principle *should* say (T) holds where x is a designation of a sentence y *which expresses one proposition*. The paradoxical sentence (a) does not express one proposition. Either it is ambiguous or, more plausibly, nonsense, expressing no proposition at all. For the supposition that it does express one proposition leads to a contradiction. This fits in perfectly with the Prior-style analysis of the paradox just given.

If you accept Prior's case that locution 1 is basic relative to locutions 2 and 3 and if you accept the idea that a sentence is true just in case it expresses one true proposition, then you ought to accept the claim that locution 1 is basic relative to locution 4.[1]

University of Nebraska-Lincoln,
1010 Oldfather Hall,
Lincoln, Nebraska 68588-0321,
U.S.A.

© CHARLES SAYWARD 1987

REFERENCE

A. N. Prior, *Objects of Thought*, edited by P. T. Geach and A. J. P. Kenny (Oxford at the Clarendon Press 1971).

---

[1] I thank Philip Hugly for helpful discussion.

QUANTIFICATION, DESCRIPTION, AND IDENTITY

By HERBERT HOCHBERG

QUINE has long maintained, in keeping with his slogan that 'to be is to be the value of a variable,' that quantification is the key to ontological commitment. He has also maintained that proper names can be systematically replaced by Russellian type definite descriptions in a perspicuous symbolism. A simple point that has been overlooked in the many debates surrounding these themes is that in one fundamental sense in which it is taken, the first assertion is a simple consequence of the second.[1] We can see that as follows.

Quine has sometimes claimed that, where 'α' is a purported proper name, we can always use the context '$x = \alpha$' to form a description '$(\imath x)\ (x = \alpha)$' which can replace 'α', if other uniquely descriptive contexts are not readily available. At other places he has claimed that we may introduce the predicate 'α-izes' as a primitive predicate (thought understood to mean 'being-α') instead of as an abbreviation of the expression '$x = \alpha$' or of '$\hat{x} = \alpha$', where the latter is, in Russell's notation, a predicate abstract, like '$(\lambda x)\ (x = \alpha)$', and not an open sentence.[2] This difference, if it is a difference, makes no difference to the present discussion.

---

[1] Another sense in which the quantifiers may be taken to involve ontological commitments is furnished by the familiar Russellian idea that quantifiers are understood to be connected with a domain of objects over which they range. This is related to Russell's rejection of impredicative functions and his theory of types. Such a connection of quantification with ontological commitment, while related to the question being discussed and to other issues posed by free logics, is not really germane to the point at hand.

Adopting such a Quinean pattern of replacement, consider universal instantiation. On a Russellian treatment of definite descriptions, the appropriate theorem of *Principia Mathematica* is *14.18:

$$E!(\iota x)\ (\Phi x) . \supset : (y)\ \Psi y . \supset . \Psi(\iota x)\ (\Phi x).$$

Thus, we would have:

$$\frac{E!(\iota x)\ \Phi x \ \&\ (y)\ \Psi y}{\therefore \Psi(\iota x)\ (\Phi x)}$$

as the corresponding universal instantiation pattern, adapting the *Principia* pattern to a natural deduction system. This means that if we accept the inference pattern:

$$\frac{(y)\ \Psi y}{\therefore \Psi(\iota x)\ (\Phi x)}$$

for a description '$(\iota x)\ (\Phi x)$', we have implicitly assumed '$E!(\iota x)\ (\Phi x)$'. And this, simply put, amounts to Quine's claim that quantification involves ontological commitment: that the acceptance of universal instantiation or existential generalization with respect to a sign commits us to the existence of an entity designated by the sign.

The point is even more obvious in the case of existential generalization, where the relevant theorem of *Principia* is

$$\Psi(\iota x)\ (\Phi x) . \supset (\exists y)\ \Psi y,$$

which is a trivial consequence of the contextual definition of definite descriptions. Thus, it is not only needless but vacuous to employ

$$\frac{\Psi(\iota x)\ (\Phi x)\ \&\ E!(\iota x)\ (\Phi x)}{\therefore (\exists y)\ \Psi y}$$

as the pattern for existential generalization with definite descriptions.

It is worth noting that the use of patterns like

$$\frac{\Psi\alpha\ \&\ (\exists x)\ (x = \alpha)}{\therefore (\exists y)\ \Psi y} \text{ and } \frac{(y)\ \Psi y\ \&\ (\exists x)\ (x = \alpha)}{\therefore \Psi\alpha}$$

---

[2] For a discussion of Quine's use of the context '$x = \alpha$' see my 'Mapping, Meaning and Metaphysics' reprinted in H. Hochberg, *Logic, Ontology and Language* (Munich: 1984), pp. 170-3.

in so-called free logics simply amounts to the application of Russell's pattern for definite descriptions to purported names (that may not name) and, hence, to the adoption of Quine's basic theme.

In his way, Quine was one of the first contemporary philosophers to advocate a so-called free logic, without explicitly calling it such.[3] In effect, Quine's insistence on quantification being the key to ontological commitment is adopted by advocates of free logics. This is ironic in that a characteristic that Russell took to be an essential difference between proper names and descriptions – that the logical patterns for names and descriptions differ in that names do not require an accompanying existential clause, in the application of universal instantiation, since to be a name is to name – is denied by both Quine and free logicians in that both apply Russell's inference pattern for definite descriptions to all singular terms. This is done by Quine when he turns all singular terms into definite descriptions: it is done by advocates of free logics when they admit names that do not name and apply the Russellian inference pattern for definite descriptions to names.

Quine's introduction of 'izing' predicates has been adopted by some philosophers, who advocate the adoption of free logics, to represent purported essential, individuating properties. Thus, we hear about 'possible worlds' which contain the property $\hat{x} = \alpha$, but not the individual α, and this supposedly jibes with the acceptance of non-naming names. There is a fundamental problem involved in such a view.

Consider the relational predicate 'R', representing the relation right-of, and the atomic sentence '$\alpha R \beta$'. By abstraction from this sentence we obtain the abstract '$\hat{x} R \hat{y}$'. A question arises as to whether we treat '$\hat{x} R \hat{y}$' as merely another expression for the relation R or distinguish, as Russell *sometimes* did, between the relation R and the propositional function $\hat{x} R \hat{y}$ (where the propositional function is not a linguistic expression but the function *represented* by the expression '$\hat{x} R \hat{y}$'). If we make the distinction, it should be clear that the expression '$\hat{x} R \beta$', obtained by partial abstraction from the atomic sentence '$\alpha R \beta$', will also represent a propositional function, if it represents anything at all. One can, if one chooses, call such a function a relational property, just as one can call $\hat{x} R \hat{y}$ a relation. But, in so far as one distinguishes R from $\hat{x} R \hat{y}$, this is misleading. If one identifies $\hat{x} R \hat{y}$ with R, then it matters little how one speaks of $\hat{x} R \beta$. Since, for our purposes, it does not matter how one construes '$\hat{x} R \hat{y}$', we shall simply take it to be another sign for the dyadic relation R and speak of $\hat{x}R\beta$ as a relational property.

---

[3] Quine was interested, quite early, in another aspect of free logics: the concern with an empty domain in validity theory. On this concern of his see my 'A Note on the Empty Universe,' *Mind* 66, 1957, pp. 544–6. It is also worth noting that some time ago Quine, casually but explicitly, discussed truth conditions for the atomic formulae of a schema which would contain proper names that did not name: see *From a Logical Point of View* (Cambridge: 1953), pp. 166–7.

Since we can form the sentence '$\alpha R\alpha$', we can also obtain '$\hat{x}R\hat{x}$' by abstraction. A question then arises as to whether we have a sign for a different dyadic relation or, perhaps, for a monadic property – the property of having R to itself. We can easily see that it will not do to take '$\hat{x}R\hat{x}$' as a monadic property. Let '$\alpha R\beta$' be true and '$\alpha R\alpha$', of course, be false. It should then be obvious that to assert (1) that α has R to β, (2) that α does not have R to itself, and (3) that β does not have R to α must be understood as claiming that exactly what is asserted to obtain between α and β is asserted not to obtain between α and itself and between β and α. What is asserted must then be that one and the same relation, the relation R, does obtain in the one case and does not obtain in the other two cases. What is asserted cannot be that one and the same monadic property obtains in the one case and does not obtain in the other two cases. Thus, it will not do to hold that '$\hat{x}R\hat{x}$' represents a monadic property that is asserted not to hold of α.

If one acknowledges $\hat{x}R\hat{x}$, one should then hold that it is a dyadic relation that does not obtain of the pair $\langle\alpha, \alpha\rangle$, as the relation R does not obtain of the pair $\langle\beta, \alpha\rangle$ (or, if one prefers, of the elements of the pair in the appropriate order). But, $\hat{x}R\hat{x}$ cannot be taken as a dyadic relation, just as it cannot be taken as a monadic property. For, it is also clear that to assert that '$\hat{x}R\hat{x}$' is not satisfied by the pair $\langle\alpha, \alpha\rangle$, is simply to assert that 'R' (or '$\hat{x}R\hat{y}$') is not satisfied by $\langle\alpha, \alpha\rangle$, just as to say that '$\hat{x}R\hat{x}$' is satisfied by $\langle\alpha, \alpha\rangle$ is to say that 'R' is satisfied by $\langle\alpha, \alpha\rangle$. It is not just a question of $\hat{x}R\hat{x}$ being a superflous relation, according to a principle of ontological economy along the lines of Occam's Razor, but that if one recognizes such a relation one must also recognize distinct situations corresponding to ascriptions of '$\hat{x}R\hat{y}$' and of '$\hat{x}R\hat{x}$'. Thus, to ascribe both relational predicates to $\langle\alpha, \alpha\rangle$ is to assert '$\alpha R\alpha$'. Yet, in one case, we purportedly ascribe one relation and, in the other case, the other relation. This is patently absurd. We can no more take $\hat{x}R\hat{x}$ to be a relation than to be a monadic property.

What holds for $\hat{x}R\hat{x}$ holds for $\hat{x} = \hat{x}$. There is no more a monadic property or relation of self-identity, in addition to the relation of identity, than there is such a property or relation of self-right-of. If this is the case, there is likewise no property of being-α. To say that β does not have the property $\hat{x} = \alpha$ is simply to say that $\langle\beta, \alpha\rangle$ does not satisfy '=' (or '$\hat{x} = \hat{y}$') – that the identity relation does not obtain between β and α.

If one recognizes the property $\hat{x} = \alpha$, i.e. being-α, then, by the same pattern, one recognizes the property of *being*-(being-α), and so on ad infinitum. Just as in the case of Fregean senses, where one may generate an infinite hierarchy of senses grounded on the sense of a particular term, so one can generate the infinite hierarchy of essential properties:

$$\hat{x} = \alpha, \hat{\varphi} = (\hat{x} = \alpha), \hat{\Phi} = (\hat{\varphi} = (\hat{x} = \alpha)), \ldots$$

grounded by the individual α. This is like the set theoretical pattern, whereby an infinite hierarchy of unit sets is generated by any object. One is thus led to suspect that the introduction of essential properties, such as $\hat{x} = \alpha$, may be due to thinking of properties along set theoretical lines, irrespective of the claim that $\hat{x} = \alpha$ can exist when α does not. There is a second connection that such essential properties have to sets if we ignore such a claim.

As mentioned earlier, one fundamental or logical difference that Russell took to distinguish definite descriptions from logically proper names was the different inference patterns that applied to the two kinds of signs. Another difference concerned the question of the scope of descriptions. Such differences may be called logical in a fairly clear sense. A basic difference between sets and properties has always been held to be given by the so-called identity conditions for sets. Where 'S', 'T', and 'C' are appropriate variables ranging over sets:

$$(S)\ (T)\ [(x)\ (x \in S \equiv x \in T) \supset S = T]$$

and hence:

$$(S)\ (T)\ [(x)\ (x \in S \equiv x \in T) \supset (\mathrm{C})\ (S \in \mathrm{C} \equiv T \in \mathrm{C})]$$

reflect the extensionality of identity for sets, and are incorporated in the logic of sets. By contrast, for properties, we do not have either:

$$(f)\ (g)\ [(x)\ (fx \equiv gx) \supset f = g]$$

or, with 'Φ' as a second level predicate variable:

$$(f)\ (g)\ [(x)\ (fx \equiv gx) \supset (\Phi)\ (\Phi f \equiv \Phi g)].$$

But, in the case of purported properties like $\hat{x} = \alpha$ and $\hat{x} = \beta$, we do have:

$$(y)\ [(\hat{x} = \alpha)\ y \equiv (\hat{x} = \beta)\ y] \supset (\hat{x} = \alpha) = (\hat{x} = \beta),$$

with the parentheses in '$(\hat{x} = \alpha)$' and '$(\hat{x} = \beta)$' serving to separate these abstracts, used as predicates in the conditional's antecedent, from the subject sign 'y' and, used as subject signs in the consequent, from the dyadic predicate '='. Hence, we also have:

$$(y)\ [(\hat{x} = \alpha)\ y \equiv (\hat{x} = \beta)\ y] \supset (\Phi)\ [\Phi(\hat{x} = \alpha) \equiv \Phi(\hat{x} = \beta)]$$

with the abstracts '$(\hat{x} = \alpha)$' and '$(\hat{x} = \beta)$' again occurring in predicate place in the antecedent and in subject place in the consequent of the conditional. (Possible complications posed by contexts with intentional verbs are ignored as being irrelevant to the issue.) The point is that the purported essential properties behave remarkably like sets in a crucial respect. Moreover, this is also true for all the properties in the infinite hierarchy of purported properties that are generated by α. This serves to cast additional suspicion on the claim that an abstract like '$\hat{x} = \alpha$' represents a property.

The University of Texas at Austin,
Waggener Hall 316, Austin,
Texas 78712-1180, U.S.A.

© HERBERT HOCHBERG 1987

MORRIS ON IDENTITY

By GEORGE ENGLEBRETSEN

CITING earlier support for his thesis from P. F. Strawson (*Subject and Predicate in Logic and Grammar*, London 1974) and M. J. Lockwood ('Identity and Reference,' in *Identity and Individuation*, M. K. Munitz, ed., New York 1971), Thomas V. Morris (*Understanding Identity Statements*, Aberdeen 1984, esp. ch. 4) has recently claimed that identity statements of the form 'A is B' must be taken as informative in the sense that they are used to collate or merge pairs of bodies of information. The idea here is that in addition to the ontological, logical and linguistic associations connected to any nonvacuous referring expression there are epistemic associations as well. Each such expression has associated with it a body of information relative to each user of that expression. To use Lockwood's vivid image, each of us has a kind of mental file for each item of our acquaintance. Suppose 'Phosphorus' is the label of one of my files. Upon learning that Phosphorus is a planet rather than a star I mentally delete from my file the claim 'Phosphorus is a star' and add the claim 'Phosphorus is a planet'. Suppose 'Hesperus' is a label on another of my files. In it I keep my beliefs concerning Hesperus. Now suppose that I learn that Phosphorus is Hesperus, i.e. that 'Phosphorus is Hesperus' is true. According to the thesis defended by Morris, the epistemic function of this statement differs from that of 'Phosphorus is a planet' (which adds to my mental file on Phosphorus), and it differs from that of 'Phosphorus is Phosphorus', which has no epistemic function at all. Instead, it, like any informative identity statement, collates or merges bodies of information. In effect, when I learn that Phosphorus is Hesperus I do learn something (unlike the 'Phosphorus is Phosphorus' case), but I do not thereby *add* anything to my mental files (unlike the 'Phosphorus is a planet' case). Rather, I simply delete one file and place all of its contents into the other. Thus I might take all the information formerly in my Hesperus file (minus the claim that Hesperus differs from Phosphorus) and place it in my Phosphorus file. While acceptance of a predication adds something to my mental file, and acceptance of a tautology neither adds to nor subtracts from it, acceptance of an identity subtracts (through merger) one entire file.

It's a nice picture. But it does not solve the problem for which it was designed. The problem was how to account for the fact that sentences like 'Phosphorus is Hesperus' are informative while those like 'Phosphorus is Phosphorus' are not. Morris had earlier criticized the view that sentences like 'A is B' are, à la Frege, about not A and B but 'A' and 'B'. Thus 'A is B' says that 'A' and 'B' name the same object. Yet this clearly generates an infinite regress since, on this analysis, 'A is B' is read as 'The object named by "A" is the object named by "B" ', a sentence having the form of the original. Now Morris's thesis suffers from the same sort of regress. 'A is B' is informative in the sense that the A and B files are merged. Yet surely one piece of information in the newly augmented file is that conveyed by the sentence 'A is B'. But what information do I now have when I know that 'A is B' is true?

Bishop's University,
Lennoxville, Quebec,
Canada J1M 1Z7

© GEORGE ENGLEBRETSEN 1987

THE PUZZLE OF THE FLASH STOCKMAN: A REPLY TO DAVID LEWIS

By PETER LAMARQUE

DAVID Lewis tells us of the song about Ugly Dave, the bragging stockman: 'I'm duke of every blasted thing . . . '.[1] According to Lewis's analysis of fiction, what is 'true in fiction' is what is true in worlds where the story is told as known fact.[2] A storyteller pretends to tell the truth about what he knows. Yet the singer of the song about Ugly Dave does no such thing. He must pretend to be Ugly Dave telling boastful lies. This raises a problem:

> When the singer pretends to be Ugly Dave pretending to tell the truth about himself, how does this differ from pretending to be Ugly Dave *really* telling the truth about himself? . . . We must distinguish pretending to pretend from really pretending. Intuitively it seems we can make this distinction, but how is it to be analyzed? (*Philosophical Papers*, Vol. I, p. 280)

---

[1] David Lewis, *Philosophical Papers*, Vol. I, Oxford 1983, p. 279.

[2] David Lewis, 'Truth in Fiction', *American Philosophical Quarterly*, XV (1978), reprinted in *Philosophical Papers*, Vol. I.

Lewis has 'no solution to offer'.

I suggest that some light can be shed on this by considering the case of mimicry or impersonation. Impersonation is a matter of imitating, and exaggerating, certain salient characteristics of the subject. We are reasonably familiar with the characteristics of (1):

(1) Ronald Reagan giving a press conference

So with a little skill we could manage (2) and (3); with rather more skill, (4):

(2) My impersonation of (1)
(3) Your impersonation of (1)
(4) My impersonation of (3)

Assuming that (2) and (3) differ in *their* salient characteristics – impersonations after all inevitably reflect the personality of the impersonator – it follows that (4) need not be identical with (2). My impersonation of your impersonation of Reagan need not collapse into my impersonation of Reagan *simpliciter*.

So it is with pretending. Or rather the kind of pretending that is at issue is a kind of impersonation. Pretending, like impersonation, has characteristics of its own which can be mimicked. The difference Lewis wants between pretending to pretend and just pretending is the difference between mimicking a pretence and mimicking the real thing.

One problem in applying this to the song might seem to be the fictionality of Ugly Dave. How can I, the singer, pretend to be Ugly Dave when there is no such person? The answer is of course that it is the *properties* characterizing Ugly Dave that I must pretend to have. There is a further problem, though, for a crucial property of Ugly Dave is his unreliability in telling us what those properties are. How do we know he is unreliable? And how do we portray it in the song?

We know it by inference. Ugly Dave doesn't tell us straight out and he is our only source of information. To discover that a fictional narrator is unreliable we draw on background knowledge. In this case we know – or believe that it is widely believed – that a genuinely outstanding stockman wouldn't adopt the manner, the tone, or the superlatives of Ugly Dave in recounting his achievements. These, on the contrary, are all the hallmarks of a boaster and a liar. We know this also from a quite different source: the conventions of the genre.

How can all this be portrayed? The singer, in pretending to be Ugly Dave, must mimic the characteristics of a boaster, that is, someone pretending to be something he is not. This is by no means the same as pretending to be Ugly Dave telling the truth about himself. The salient characteristics of a lying boaster are different from the salient characteristics of a truth-telling high-achiever.

Such, I suggest, is a basis for the analysis that Lewis is looking for. One or two more general observations come out of it. First of all, in the example, the singer's 'pretence' is a kind of mimicry but it is implausible to generalize this notion of pretence to all storytelling. Tolstoy was no mimic in writing *War and Peace*. Second, and more positively, the example reveals the close connection between character and characteristics. This gives a valuable hint about ontology. There is nothing more to Ugly Dave – unlike Ronald Reagan – than the characteristics we mimic. Yet, lastly, it seems that to some extent the singer, by adopting a certain tone, can determine what those characteristics are, perhaps even, at the extreme, turning Ugly Dave into an honest man. Some, if not all, the characteristics of Ugly Dave are *relative to a version on interpretation* of the song.

These observations suggest several inadequacies in Lewis's possible world analysis of fiction. As he admits, the unreliable narrator poses a problem or his account of what is 'true in fiction'. There is not a unitary kind of pretence in storytelling (indeed at times there is no pretence at all). Worse still, what is true in fiction, it seems, depends at least partly on certain 'intentional' factors, such as how the fiction is interpreted. Furthermore, characters are perhaps better thought of not as individuals in possible non-actual worlds – that makes them too determinate – but as kinds or characteristics in this world subject to interpretative variation. Your rendering of Ugly Dave, like your singing, might be a great improvement on mine.

University of Stirling,
Stirling FK9 4LA

© PETER LAMARQUE 1987

THOUGHTS AND DEFINITIONS

By JUSTIN BROACKES

I

IN the *Grundlagen*, Frege seems to hold that the two sentences

(1) *a* is parallel to *b*

and

(2) The direction of *a* = the direction of *b*

(and correspondingly, the two sentences 'the F's can be one-to-one correlated with the G's' and 'the number of F's = the number of G's') have the same content. In using (1) to give a contextual definition of the direction terms in (2), 'We carve up the content in a way different from the original way, and this yields a new concept.' (§64) The idea seems to be that (1) and (2) express the same content, only in different linguistic form. Frege had not articulated the distinction between sense and reference in 1884; but if the 'content' of a sentence here corresponds to what he later called its *sense*, then Frege is implying that (1) and (2) have the same sense. And though one must be cautious about interpreting old ideas in new language, Frege came close to saying precisely this in the recapitulation at the end of the book, when he says that his task has been this: 'to fix the sense of a numerical identity, that is, to express the sense [*den Sinn ... auszudrücken*], without making use of number words or the word "number" ' (§106).

The view that (1) and (2) are synonymous accords well with what Gareth Evans called 'the single constraint Frege imposed upon his notion of thought': the 'Intuitive Criterion of Difference', namely that 'the thought associated with one sentence S as its sense must be different from the thought associated with another sentence S' as *its* sense, if it is possible for someone to understand both sentences at a given time while coherently taking different attitudes towards them, i.e. accepting (rejecting) one while rejecting (accepting), or being agnostic about, the other.' (v. Gareth Evans, *The Varieties of Reference*, ed. John McDowell, Oxford 1982, pp. 18–19.) But this criterion, though it is extremely helpful in fixing Frege's notion of a thought, gives at best a sufficient condition for two sentences to express different thoughts, and not a necessary one. The criterion concentrates on the attitude of belief. But discernibility with respect to that attitude cannot be necessary for non-identity of thoughts – for it is impossible to understand '$2 + 2 = 4$' and '$2 + 1 = 3$' without believing them both, though it would surely be unsatisfactory to say that they have the same sense. It would not seem to solve the problem, simply to broaden the range of attitudes mentioned in the criterion to include desire, hope, intention or other familiar attitudes.

We might try to broaden the range of 'attitudes' to include inferential relations – on the ground that the thought that $2 + 1 = 3$ and the thought that $2 + 2 = 4$ could be distinguished, since the first, taken with the additional premises that $1 + 1 = 2$, that $3 + 1 = 4$ and that if equals are added to equals then the results are equal, implies that $2 + 2 = 4$, whereas the latter implies it directly and trivially. The trouble with this suggestion is that, though it is definitely true that if thoughts a and b have different inferential relations then they are different, this criterion will typically be inapplicable in precisely the cases where guidance is needed. For *are* the inferential relations of (the thought expressed by) 'The direction of a = the direction of b' the same as those of (the thought expressed by) 'a is parallel to b'? Someone who thinks they express different thoughts will say that the former, together with 'the direction of a = 35 °E', directly implies 'the direction of b = 35 °E', whereas the latter does not. But someone who thinks they express the same thought will simply deny this, insisting that the latter sentence has just the same direct implications as the former – though a formal proof might well set out the two supposedly synonymous sentences on separate lines. And it is not easy to see a way of deciding between these rivals.

The way out of these difficulties is to recognize a *further* ingredient in the notion of a thought, besides the fact that thoughts are the objects of attitudes and stand in inferential relations. Thoughts are structured, and their elements are concepts: having any thought is the exercise of a *complex* ability, aspects of which are individual conceptual abilities, each of which is capable of being employed also on other occasions. Frege expresses this idea graphically: 'thoughts are constructed out of building-blocks. And these building-blocks correspond to groups of sounds out of which the sentence which expresses the thought is built, so that the construction of the sentence out of its parts corresponds to the construction of the thought out of its parts.' (v. 'Logik in der Mathematik', 1914, in G. Frege, *Nachgelassene Schriften*, ed. H. Hermes, Fr. Kambartel, Fr. Kaulbach, Hamburg, 1969 p. 243; and in *Posthumous Writings*, tr. P. Long, R. White, Oxford 1979, p. 225.) Evans develops the idea clearly in his 'Generality Constraint': 'if a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception.' (v. *The Varieties of Reference* p. 104.)

This kind of consideration permits us to distinguish the thought expressed by 'the direction of a = the direction of b' from that expressed by 'a is parallel to b'. For the former employs a conceptual ability which is also used in having the thought that the direction of a is not a steam-roller – and the latter does not. Someone who understood ' – is parallel to – ' but had not yet mastered the concept of a direction could not have the thought that the direction of a is not a steam-roller. And if the conceptual abilities employed in understanding the sentence 'a is parallel to b' do not change when a speaker acquires the additional ability to think about directions, then the sense of that sentence must remain different from that of 'the direction of a = the direction of b', even for someone who understands them both and understands their necessary equivalence. When using the former sentence he will be employing various abilities that are reused on other occasions; when using the latter sentence he will be employing different abilities which are reused on a different set of other occasions. The only way Frege could, I think, maintain the view that the two sentences are synonymous would be to say that it makes no sense to ask – of someone who has *both* ranges of conceptual abilities – *which* range is being employed when he produces one of our two sentences, though it would make sense when he employed a sentence that appeared in one discourse without having an equivalent in the other. But I can see no call for such scepticism about the notion of conceptual ability.

The 'intuitive criterion of difference' – the possibility of taking different epistemic attitudes to *a* and *b* – provides a sufficient criterion for two thoughts to be different, but not a necessary one. There is a further condition, that thoughts will be different if they involve different conceptual abilities – which implies that contextual definitions, if they introduce new concepts, will never provide mere synonyms for sentences involving that concept. It is perhaps strange that Evans, who, with his Generality Constraint, was in such a good position to see this, should yet have presented the former condition as 'the single constraint Frege imposed upon his notion of thought.'

II

This raises an important question: if sentences employing a new concept cannot be synonymous with sentences employing old concepts, then how can contextual definition ever succeed in introducing someone to a new concept? In a sense, I think a good contextual definition need not be capable of *always* doing so: precisely because direction-discourse involves new conceptual abilities, beyond those available to a 'primitive' who knows only how to talk about lines and parallelism, it is always a possibility that when we give the primitive a contextual definition, he may be unable to follow it. Stretched to the limit of his conceptual ability, he may fail to catch on: he may at best manage laboriously to transform direction-sentences into their equivalents – and that is not the ability possessed by someone equipped with an ordinary mastery of direction-discourse.

How then does a contextual definition *ever* work? I think it works by providing a recipe for conceptual innovation, giving a rule for finding a sentence in the old discourse that is *a priori* equivalent to (though not synonymous with) any sentence in the new. Given the recipe, some people may cotton on to the new concept, others may not. But it is a mistake to demand more from a contextual definition: precisely because a *new* conceptual ability is in question, the old abilities will not be sufficient in themselves to provide it. At best therefore, we can provide pointers and training; but we cannot assume that the subject will catch on.

This provides Frege's contextual definitions with a means of escape from a suspicion that Dummett has raised, that such definitions may be incompatible with Frege's realism about abstract objects. If we had a contextual definition of direction-discourse,

> we could then no longer claim that the sense of a term for a direction was given as a means of determining its referent, or that the determination of the truth-value of a sentence containing the term went via the referent. On the contrary, the most direct means of establishing such a sentence as true or false, that which corresponds, step by step, with the way in which the sense of the sentence is given, will consist in first transforming the sentence, in accordance with the contextual definition, into one about lines, and then establishing the truth-value of the latter sentence; in this way, the referent of the direction-term is by-passed altogether .... If the stipulation is carried out in a logically unobjectionable manner, then no absurdity can arise from crediting those names with the property of standing for objects. It nevertheless remains that the notion of reference is not really *used* in the explanation of the senses of the terms in question .... Numerical terms, if explained in such a way, would, as it were, be said to have a reference only by courtesy.
>
> Viewed in this light, Frege's realism concerning abstract abjects appears to reduce to little more than a form of words. (M. A. E. Dummett, *The Interpretation of Frege's Philosophy*, London 1981, pp. 425-26.)

This would be a fair criticism if sentences containing the definiendum were *synonymous* with sentences containing the definiens: the attribution of reference to the definiendum would indeed be idle, as far as the assessment of the truth-value of the former sentences is concerned. But I have argued that the former sentences will not be synonymous with the latter: someone who properly cottons on will develop ways of using the former other than as mere abbreviations for the latter, and with this will come ways of assessing the truth-value of the former sentences other than via their *a priori* equivalents. A contextual definition will not have fully introduced someone to direction-discourse until he has seen how the sense of a direction-term can be taken as a route to a reference, namely, a direction.

If this is right, however, Frege's admission of abstract objects like numbers and directions has a slightly different justification from that which is commonly supposed. It rests not directly upon the 'context principle' and the provision of contextual definitions for numerals and direction-terms, but rather upon the existence of abstract names, and the fact that the notion of an object is correlative with that of a name: we have no notion of an object other than that of the reference of an actual or possible name. (This correlation of names and objects is clearly defended in Chapter 4 of Dummett's *Frege: Philosophy of Language* (London, 1973, 1981).) The principle that only in the context of a sentence does a word have a meaning seems to me (on several natural readings) both interesting and true; but it is the *use* of direction-terms *as names* (in certain appropriate ranges of sentences) that ensures that we are right to say that they name objects, not the availability of transformations of such sentences into *a priori* equivalents in a conceptually simpler discourse.

Oriel College, Oxford © JUSTIN BROACKES 1987

HOW MANY MINDS HAS A TREDECIM?

By GEORGE REA

A TREDECIM is a philosophically interesting fictional beast devised by Hugh Chandler in 'Indeterminate People', ANALYSIS 45.3, June 1985. A tredecim is the temporary composite form of a thirteen species symbiotic cluster. It is difficult to imagine what shapes of thirteen different species might make up a suitable combination, but we do not need to concern ourselves with this problem. Chandler stipulates that the tredecims have a language of their own, and they are capable of more complex action than any of their components, for instance housebuilding and musical composition. After disintegration the components say they remember having been engaged in difficult 'computational tasks'. Suppose that these apparent memories are authentic. We may well wonder what were the components' motives for performing these difficult tasks. Presumably the individual components did not each understand the method and reaons for housebuilding and how their individual tasks fitted in. More likely, they did not understand their tasks, but they simply felt an instinctive need to perform, perhaps with a vague goal of shelter or some other pleasure. But do not these 'instinctive needs', because of the complexity of the collective task, themselves require personal explanation in terms of planning and intentional stimulus? These are points Chandler does not develop.

Furthermore, significantly, Chandler has also left open the possibility that the components' apparent memories of tredecim life are not authentic. We are only told what they seem to remember. Chandler says nothing definite about the components' mental lives within the tredecim, except for the negative point that when the tredecim cries on disintegration, no component feels pain. The question is then posed whether the disintegration is painful. This is related to the question whether the tredecim is a person or a team. However Chandler does not make clear what he means by this. In the end he does not decide either way on this question and suggests that, since we could decide either way, we are compelled to be non-realists in regard to the meaning of the tredecim's cry. But Chandler's analysis does not convince me that there is any indeterminacy in this matter which is not mere fictional indeterminacy (i.e. author's indecision). Chandler has not analysed what sort of minds might exist in a tredecim. I intend to show that real differences within the mental life of a tredecim would determine the answers to Chandler's questions.

First we should take note of Richard Swinburne's point that for a mind M to belong to a body B is for B to be the vehicle of the agency and (more important) the experience of M (S. Shoemaker and R. Swinburne, *Personal Identity*, Blackwell 1984, p. 22).

The first sort of tredecim I shall describe is a single-minded person-like tredecim, T1. When a T1 cluster amalgamates the nervous systems of the components interconnect. The individual minds of the components cease to exist as distinct entities. That is to say, for instance, that when T1 amalgamates, there is no longer a mind which sees through only the eyes of component A or which feels pains in only part A of the tredecim body. In this sort of sense we would say that the subsequent apparent memories of the components concerning tredecim life are in this case not reflections of components' earlier experience, but are merely after effects of tredecim life.

Suppose each component has two eyes. There is a mind MT1 which has a much broader scope of vision than any of the components. MT1 sees through all twenty six eyes of T1. Likewise MT1 has other sensations such as hearing and touch from all the other sensory apparatus within the cluster. And MT1 has direct control (by basic actions) over all the limbs of the cluster.

MT1 has access to a range of memory-like, mainly true, beliefs about the past experiences and thoughts of each of the components. Just as you can intentionally have a memory image of a particular vision and believe that it is a representation of a vision which you had in the past, so MT1 can summon up an image as of a past vision and believe (correctly) that it is a representation of the vision which component A had at t. MT1 can also have a different image of another past vision and believe that it represents a vision of component B also at t. This also holds for similar memory-like belief about other past experience and thoughts of the components. We may call such memory-like phenomena, in any sort of tredecim, 'T1 memory', prefixing the verb 'remember' accordingly. Thus MT1 can T1-remember various separate past experiences and thoughts of all the components. Despite access to all the component languages, T1 type tredecims have developed their own language.

T1 satisfies the criteria of person-likeness which Chandler indicates, in that T1 has a mind which understands how (and why) to build houses and compose music, and T1 feels pain on disintegration.

Let us try to make sense of the notion of a fourteen-mind tredecim, T2. In T2 the components' nervous systems do also interconnect, but in a different way from T1. In T2 there is a mind MT2 like MT1 in respect of experience and T1-memory. But the movements of T2 have a degree of independence of the desires of MT2.

The minds of the components continue their individual existence in T2, i.e. there is still a mind which sees through only the eyes of component A and which also controls only the limbs of part A of T2, and so on. However their mental lives do of course undergo significant changes. The intentions of MT2 automatically stimulate desires to perform relevant actions within the components. The components find such desires beneficial for the tredecim as a group (as well as for MT2), and they do each feel a genuine concern for the rest of the group as well as each for itself, so that they usually follow up their impulses. MT2 does all the more complex thinking, while the components are generally happily subservient. MT2 feels pain on disintegration.

We can also imagine a thirteen mind simple 'team' tredecim, T3. However T3 offers less explanatory power regarding the complex action of the tredecim. T3 is just like T2 but without the existence of a transcendent fourteenth mind like MT2. The disintegration of T3A is not painful. But perhaps it is possible that there be a T3 tredecim, T3B, such that on disintegration, although there is, for instance, no mind which sees through component A's eyes and also feels pain (i.e. no component feels pain), there is briefly a very simple mind which has only the experience of feeling pain from all T3B's vital links during disintegration. But the brief existence of a fourteenth mind, as in disintegration of T3B, seems much less likely than that simply the components feel pain (T3C).

I have described some alternative forms of tredecim which are left open by Chandler's fictional indeterminacy. Perhaps Chandler would like to describe some further possibility. But I cannot see how any coherent account of tredecim life could amount to a case of non-realism.

8, Highfield Road,
Cleator Moor, Cumbria

© GEORGE REA 1987

KNOWLEDGE, BELIEF AND EXISTENCE

By C. J. F. WILLIAMS

IN my book *What is Existence?*,[1] I attempted an analysis of the proposition 'President Ford does not know that David Pears exists' which does not involve categorizing 'exist' as a first-level predicable, i.e., as something which by itself can say something true or false of an individual, such as David Pears. Reviewing my book,[2] Thomas P. Flint claims to have produced a refutation of this analysis. He describes a state of affairs in which we should be prepared to accept as true

(2) For some θ, both David Pears alone θ's and President Reagan believes that for just one x, θx,

while rejecting as false

(1) President Reagan believes that David Pears exists.

The circumstances envisaged are those in which David Pears, and no one else, is in the White House and no one at all is in the Capitol, whilst President Reagan believes, mistakenly, that Howard Baker alone is in the Capitol and that no one at all is in the White House. Thus

(3) David Pears alone is-either-in-the-White-House-or-in-the-Capitol and President Reagan believes that for just one x, x alone is-either-in-the-White-House-or-in-the-Capitol

is true, and since (2) is an existential generalization of (3), (2) is true also. But in these circumstances there is no guarantee that (1) is true.

I have no wish to dispute Professor Flint's claim that the truth of (2) is compatible with the falsehood of (1); but since my own concern was with a president's knowledge, not with his beliefs, I must regard his reasoning as a case of *ignoratio elenchi.* He may wish to argue *a pari* that if

(4) President Reagan knows that David Pears exists
can be paraphrased by
(5) For some θ, both David Pears alone θ's and President Reagan knows that for just one x, θx,

(1) should also be capable of being paraphrased by (2), which clearly it is not. If he does, *nego paritatem.* I did indeed argue that the negation of (5) – give or take a president or two, and with

---

[1] Oxford, The Clarendon Press, 1981, ch. IV. My argument was against J. L. Mackie, to whom is due the example about President Ford and David Pears.

[2] In *The Philosophical Review*, Vol. XCIII (1984), pp. 131-4. Professor Flint provides a lucid and accurate summary of the main theses of the book, for which I am grateful.

certain qualifications about possible substitutions for the predicate-variable which Professor Flint fails to notice – provides an analysis of the negation of (4). In the scenario suggested by Flint, however, (5) is no more likely to be true than (4). Since the President's belief about Howard Baker is, *ex hypothesi*, mistaken, we have good Gettier-type reasons for denying that he knows that for just one x, x is-either-in-the-White-House-or-in-the-Capitol. So Flint has done nothing to disprove the material equivalence, at least, of (4) and (5), nor, consequently, of their negations.

Flint's unwarranted substitution of 'believe' for 'know' prompts some remarks about the difference between knowledge and belief in this context which, perhaps, have an interest that goes beyond the question whether a particular reviewer's criticism of an author is justified. For a start, my intuitions would prompt me to give an analysis of (1), which ascribes a true belief to President Reagan, on the same lines as

(6) President Reagan believes that J. R. Ewing exists

or

(7) President Reagan believes that Bertrand Russell exists,

which ascribe false beliefs to him. In each case, an analysis is possible which, unlike that which I gave of (4), regards the words following 'that' as a proposition genuinely embedded in a wider context. The proposition 'J. R. Ewing exists' is a proposition, not about a non-existent individual, but about the status – fictional or genuine – of a proper name. (I argue this in ch. X of my book.) If (1) is understood by analogy with (6), it will similarly ascribe to Reagan a belief – a true one, this time – about the status of the name 'David Pears'. The proposition 'Bertrand Russell exists', on the other hand, is interpreted as the contradictory of the proposition 'Bertrand Russell no longer exists', and is genuinely about the old campaigner for whom, no doubt, President Reagan feels unqualified admiration. If (1) is understood by analogy with (7), it will ascribe to Reagan the true belief that David Pears still exists. (I argue in ch. V of my book that, although 'exist' is never a predicate of individuals, 'still exists' often is.) Either of these interpretations of (1) seems plausible to me; but I should not be the least bit tempted by (2), which Professor Flint wishes to foist on me.

The reason for this asymmetry between my interpretation of (4) and my interpretation of (1) is to be sought in the connection between knowing and telling and in the so-called 'factive' character of 'know'. I pointed out in my discussion of 'President Ford does not know that David Pears exists' in chapter IV of my book that we cannot *tell* President Ford what it is said that he does not know by simply saying to him 'David Pears exists'. If he doesn't know that David Pears exists, he will not know who we are talking about when we use the name 'David Pears'. Understanding another person's use of a name *as a name* requires knowledge of who its bearer is, and *a fortiori* the knowledge that it has a bearer. To inform the President of David Pears's existence, it would be no use using his name, although it may be necessary to mention it. Rather, we should tell Mr Ford something of this sort: 'There is a man called "David Pears" who teaches philosophy at Oxford, and who has written books about philosophers called "Russell" and "Wittgenstein", and who ... ' There need not be any particular set of properties unique to David Pears whose instantiation is what Mr Ford needs to know about if he is to know that David Pears exists; but if he does not know that David Pears exists, there is no such set of properties which he knows to be instantiated. What he is ignorant of is the instantiation of any such set, since we may hope to cure his ignorance just by telling him of the instantiation of some such set.

Suppose that what President Reagan is ignorant of is the fact that J. R. Ewing doesn't exist. To say that *a* does not know that *p*, just as much as to say that *a* knows that *p*, is to presuppose that it is true that *p*. (This presupposition may be cancelled by stressing the word 'know'.) Saying that Reagan either knows or does not know that David Pears exists involves the presupposition that the name 'David Pears' has a bearer. This is why it is natural to interpret 'David Pears', as it occurs in (4), as a proper name being used as a proper name. But

(8) President Reagan does not know that J. R. Ewing does not exist

carries the contrary presupposition; we have to interpret 'J. R. Ewing does not exist' as saying something true, and it can do this only if we suppose it to be about a name, not about a person. Again, we can find out something about the structure of (8) by considering the connection between *knowing* and *telling*. What would we need to tell the President if we were to try to relieve his ignorance about J. R.? 'Mr President, "J. R. Ewing" is not the name of a real oil-baron; it is just the name of a character in a soap opera. (That is to say, millions of people spend an hour a week pretending it is a proper name.)'

J. R. Ewing never has existed. Bertrand Russell once did. With

(8a) President Reagan does not know that Russell does not (now) exist

the information needed is different. 'Russell' can be used as a proper name. (Indeed, *pace* Russell, 'Russell' can be used as a logically proper name.) What we should need to tell Mr Reagan if (8a) were true is that Russell no longer exists. How this is to be analysed is something I discuss at length in chapter V of my book; but since there is no difficulty about our telling anyone that Russell does not now exist, the considerations that led to my analysing (4) as (5) do not arise.

Consider now

(9) President Reagan does not know whether David Pears exists (or not).

This carries no presupposition of fact.[3] On the other hand, it presupposes a question. It envisages the President asking himself 'Does David Pears exist?' The sense which 'David Pears exists' has in (9) must be the same as it has in the direct question 'Is it the case that David Pears exists?' The sense of a question which requires the answer 'Yes' or 'No' cannot presuppose the truth of either the affirmative or the negative answer to it. So 'David Pears' here cannot be being used as a proper name.

I have just said that (9) 'presupposes a question'. So does (1); but it behoves me to explain what I mean by this mysterious phrase. I mean this. If we say of *x* that he does not know whether *p* (or not), it must be intelligible to suppose that he should ask the question 'Is it the case that *p*?', and the words that we substitute for this last occurrence of *p* have the same sense as they had when substituted for the previous occurrence of *p*. Since 'David Pears' cannot be being used as a proper name in the question 'Is it the case that David Pears exists?', it cannot be being used as a proper name in propositions like (1), which involve belief. For, if we say of *x* that he believes (does not believe) that *p*, it must be intelligible to suppose that he should ask the question 'Is it the case that *p*?'; and the words we substitute for this last occurrence of *p* must have the same sense as they had when substituted for the previous occurrence of *p*. What I believe or disbelieve (or do not know whether ...) is precisely the same as what I wonder or ask about.

What I know is different. (4) presupposes a fact, by which I mean to indicate the familiar feature of knowing which Vendler signifies by calling 'know' a 'factive verb'. This means that alongside any assertion of the form '*x* knows that *p*' or '*x* does not know that *p*' there is the implicit assertion that *p*. And this makes it possible to anchor the content of the knowing in the relevant fact. The metaphor of anchoring can be explained thus: if I say

(10) *x* knows where she works,

I fix the content of the appropriate 'that-clause' by tying it to her actual place of work. Relative and interrogative pronouns in natural languages do similar work to quantifiers and bound variables in formal languages. Using '*l*' as a variable whose substitution instances are locatives (i.e., answers to the question 'Where?': e.g., 'in the Town Hall,' 'at Sainsbury's'), we can represent (10) as

---

[3] Cf. Zeno Vendler, 'Telling the Facts' in *Contemporary Perspectives in the Philosophy of Language*, edd. Peter A. French, Theodore E. Ueling, Jr. and Howard K. Wettstein (University of Minnesota, 1979), p. 229: ' "He did not say where she works." Why is this correct? I suspect that it is because the *wh* after a negation masks a "whether ... or ... " rather than a "that ... ": " ... did not say whether she works at *A* or at *B* or ... " And this is not a factive clause.'

(11) For some *l*, *x* knows that she works *l*.

What I would specify directly by saying '*x* knows that she works at the Town Hall' I specify indirectly by saying (10). Just so, 'Edith married the man whom Phyllis loved' specifies indirectly the fact directly specified by saying 'Edith married Arthur', if indeed Arthur was the man whom Phyllis loved.

It has often been remarked that, while we can say (10), the parallel for belief, '*x* believes where she works', is unacceptable. *Believing* does not take wh-nominals. It is, however, over-hasty to conclude from this that 'believe' cannot ever be construed with wh-forms. We cannot say '*x* believes where she works', but we can say

(12) *x* believes that she works where you believe she works.

Similarly, we cannot say 'He believes *what A* is'; but we can say 'He believes that *A* is what the Church says that *A* is'. Believing may not take wh-nominals, but it does take wh-adverbials and wh-adjectivals.

Here again the relative clause expresses what can be expressed more formally by quantifiers and variables. 'He believes that *A* is what the Church says that *A* is' can be rendered 'For some *F*, he believes that *A* is *F*, and the Church says that *A* is *F*'. Similarly, (12) is equivalent to

(13) For some *l*, *x* believes that she works *l*, and you believe that she works *l*.

There is no obligation to specify directly her place of work or *A*'s nature; we can always specify them indirectly by tying them to what someone else believes about her or to what something else says about *A*. Now what happens when I say (10) is exactly the same. Because knowing presupposes a fact, it makes no difference whether I say (11) or

(14) For some *1*, she works *1*, and *x* knows that she works *1*.

What the former contains implicitly the latter makes explicit.

It is not possible to give all propositions containing wh-clauses a quantificational analysis in this way. We cannot regard

(15) *x* wonders where she works

as the existential generalization of '*x* wonders that she works in the Town Hall', nor can any other plausible candidate be found to act as its existential instantiation. The 'where' in (12) would traditionally be regarded as a relative pronoun, and that in (10) as an interrogative pronoun. (10) and (15) would alike be held to contain indirect questions, whereas the 'where' in (12) would be thought of as introducing a relative clause. My suggestion, that (10) be assimilated to (12), on the grounds that both are existential generalizations, cuts across the traditional distinction between relative clauses and indirect questions.

Is this realignment defensible? Those who wonder where someone works, or whether *p*, can be thought of as asking themselves where that person works, or whether *p* (*je me demande* ...). The indirect questions introduced by 'wonder', 'ask', 'enquire', etc., are immediate transformations of sentences containing direct questions: 'He asked "Where does she work?" '. But (10) requires something a little more elaborate if it is to be given the indirect question analysis. '*x* knows "Where does she work?" ' will not do. It would have to be expanded into something like '*x* knows the answer to the question "Where does she work?" '. Should we then give this, rather than (14), as the analysis of (10)? On either analysis it will have to be supposed to be obtained from an underlying sentence by a process of deletion: we shall have a choice between '*x* knows that she works where she does work' and '*x* knows the answer to the question "Where does she work?" ' as the expanded version of (10).

In fact the mention of *question* and *answer* in the latter attempt at an elucidation of (10) appears to be redundant. Commenting on the form 'He knows the *answer* to the question whether *p* (the question what *ϕ*s)', Prior wrote:

> We assume here a basic form built up from an indicative and an interrogative clause, namely 'That *p* is the answer to the question q', where '*p*' stands for an indicative sentence and 'q' for a question-clause like 'whether *p*' or 'for what *x*, *ϕx*'. 'He knows q' then expands to 'For some *p*, (*a*) he knows that *p*, and (*b*) that *p* is the answer to the question q'. An adequate 'erotetic logic' will contain various theses about the question-and-answer relationship, e.g. 'For some *y*, that *y ϕ*s is the answer to the question for what *x*, *x ϕ*s', and 'Either the answer to the question whether *p* is that *p*, or the answer to the question whether *p* is that not *p*' (A. N. Prior. *Objects of Thought*, Oxford: at the Clarendon Press, 1971, p. 76).

On these principles, '*x* knows the answer to the question "Where does she work?" ' amounts to 'For some *l*, (*a*) *x* knows that she works *l* and (*b*) that she works *l* is the answer to the question "For what *l* is it the case that she works *l*?" '. Here the second clause (*b*) is trivially implied by the thesis of erotetic logic 'For some *l*, that *x ϕ*s *l* is the answer to the question where *x ϕ*s'. It is therefore redundant, and '*x* knows the answer to the question "where does she work?" ', as an analysis of (10), boils down to the same thing as (11), and this is equivalent to (14), which has the same form as (12). The indirect question analysis of (10), when fully worked out, does not seem to be in competition with the analysis which assimilates (10) to (12).

The often repeated claim that belief will not admit *wh*-forms whilst knowledge will does not properly represent the facts. Belief does admit *wh*-forms, but only where the *wh*-clause serves to link two clauses, clauses which in the formal language would appear as open sentences. The *wh*-form binds clauses together (a main clause and a subordinate clause) in the same way that two open sentences are linked by conjunction and a single quantifier binding a variable which is repeated in each sentence. The difference between knowledge and belief in this respect is that with believing the two clauses have to be spelled out, whereas with knowing the one can be implicit in the other. When I say that *x* believes she works where you believe she does, I anchor *x*'s view about her place of work in your view about the matter. When Newman says that he believes that Christ is what the Church says He is, he anchors his belief in the Church's teaching about Christ. When we say that *x* knows where she works, we anchor *x*'s knowledge in the facts. We might have said, pleonastically, '*x* knows that she works where she does work'.

What has this to do with knowing that David Pears exists? In *What is Existence*? (pp. 92. sq.) I took the liberty of stretching the sense of Russell's technical term 'primary occurrence', in such a way that I could claim that a name like 'David Pears' has primary occurrence in a proposition like (4). The justification for this, which I shall not now rehearse in detail, is that the analysis of propositions like (4), which I believe is given by propositions like (5), has the proper name occurring in one conjunct of a conjunctive open sentence where each conjunct contains a variable bound by a quantifier which lies outside the conjunctive open sentence, and where a sentential operator governs the conjunct in which the name does not occur. The name falls outside the scope of the sentential operator but within the scope of the quantifier. Now this pattern is discernible also in

(10) For some *l*, both she works *l*, and *x* knows that she works *l*.

We might say that the quasi-definite description 'where she works' has primary occurrence in '*x* knows where she works'. Propositions about knowledge are capable of this sort of analysis, i.e., are likely contexts for expressions having primary occurrence, *because they are anchored in facts*. It is because 'know' is a factive verb that (5) is plausible as an analysis of (4), and it is because 'believe' is a non-factive verb that (2) is not plausible as an analysis of (1). An analysis of belief-ascribing propositions must work for false beliefs as well as for true ones. If (2) were to work for (1) it would have to work for (6) as well. But unless we know that the beliefs ascribed in President Reagan in (1) and (6) are true, we have no justification in supposing the 'David Pears' or 'J. R. Ewing' can be used as proper names, and no reason to suppose that they can have 'primary occurrence', occurring in a conjunct of a quantified matrix which falls outside the scope of the relevant verb of propositional attitude. Where belief is concerned there are no guaranteed facts for the content clause to be anchored in; there are only questions to be presupposed. And for seemingly existential propositions substituted for *p* in 'Is it the case that *p*?' other analyses were possible.[4]

9 Woodland Road,
Bristol BS8 1TB

© C. J. F. WILLIAMS 1987

---

[4] I am most grateful to Michael Welbourne for helpful comments at various stages in the composition of these remarks.

EVIDENTIAL INSATIABILITY

By N. M. L. NATHAN

SUPPOSE that evidential justification is regressive: in cases when what justifies you in believing a proposition is that there is a further proposition which you believe and believe to be evidence for it, you must also be justified in believing this further proposition. Suppose also that in order to become justified in believing any proposition, you must actively investigate its truth-value. Suppose finally that nobody has time to conduct an infinite number of such investigations. It then follows that you are not justified in believing any proposition unless there is some proposition which you are non-evidentially justified in believing, i.e., justified in believing regardless of whether there is any further proposition which you believe and believe to be evidence for it. But what could this non-evidential justification be? Perhaps it would do simply to look at that part of the world which the justificandum purports to describe. But when you think about why such looking should justify you in believing something, it starts to seem as though your justification, if genuine, would be evidential after all. You wouldn't, it seems, be justified in believing the proposition which the looking was supposed to justify unless you believed an actual proposition about how things looked to you, and believed that this proposition was *evidence* for the justificandum. So it seems that all justification becomes evidential, when you are conscious of what the justificatory process amounts to. But then, on our initial assumptions, there can be no self-conscious justified believing at all.

How interesting is this perennial sceptical train of thought? It is interesting only if it is disturbing, and disturbing only if the active, self-conscious, evidentially regressive justified belief which it seems to show that we cannot have really is a kind of justified belief which it is reasonable to want. So far, regressive sceptics have not succeeded in showing that there is here a reasonable object of desire. They have neglected the question of what we actually want justified belief for. In what follows I distinguish three possible functions of justified belief. The first function, which I call the satisfaction of local curiosity, can, I argue, be served by a justified belief which requires no regressive steps on the believer's part, and the second, or stabilizing function, can be served by a justified belief which requires the believer to make only a manageable number of regressive steps. But the third function, of banishing global suspicion, can I think be served only by the kind of justified belief which, according to the initial sceptical train of thought, we can never have. This is what makes regressive scepticism disturbing.

I LOCAL UNCERTAINTY

Suppose that you wonder whether or not some particular proposition is true, and want justification, either for believing that it is true or for believing that it is false, simply as a way of making up your mind. You don't believe that circumstances will simply present you with a belief as to its truth-value, so you look for evidence, or look directly at what the proposition purports to describe. Now suppose also that the only reason why you want justified belief as to the truth-value of *any* proposition is that you want to make up your mind as to its truth-value. Will you in these circumstances, and with this attitude, have a way which satisfies you of bringing justificatory regression to a halt? Obviously you will, if there is enough about which your mind is already made up. Suppose you adduce what you take to be evidence for the proposition p whose truth-value you began by being uncertain about. Or, initially attempting a non-inferential mode of justification, say mere looking, you recognize that your continued justification is contingent on there being evidence for p, say in the shape of a proposition about how things look. Let q be the proposition you adduce as evidence for p or come in this last way to think is evidence for p. If you do thus accept q as evidence, and if your *only* reason for wanting justified belief as to the truth-value of any proposition is that you are uncertain as to its truth-value and want to make up your mind, then you have no reason for wanting to carry the justificatory process any further. For if you accept q as evidence you believe that q is true, and so far as q is concerned, your mind is already made up. No justificatory regress is able to get underway.

II THE STABILITY OF BELIEF

The first reason for wanting justification was simply to make up one's mind about whether a particular proposition was true or false. But you might also want justification for the sake of causally reinforcing an existing belief. You may be afraid that wishful thinking or illusion will eventually stop you believing what you now believe to be true. If you can now find justification for your existing belief, and can later remember the justification you now find, then your belief will be more secure than it would be if, when your belief comes under pressure, you have no evidence to refer to. But now one can imagine a regress. Suppose you do now have evidence for a proposition *p* which you already believe: either you looked for evidence or, reflecting on an initially non-inferential justification, you recognized that its validity depends on your having evidence. If *q* is this evidence, then *q*, as we have seen, will also be a proposition which you now believe. But now you can worry about the stability of your belief that *q*, and thus want causally reinforcing justification for that belief in turn. And so on.

This particular regress of justification may not however be too difficult to terminate. You believe that if you do not have evidence *q* for *p*, then your belief that *p* will be unstable, and that if you do not have evidence *q*1 for *q* then your beliefs that *q* and that *p* will be unstable, and that if you do not have evidence *q*2 for *q*1 then . . . , and so on. It is only necessary for there to be a proposition *qn* which you believe to be evidence for *qn*-1 and which you think does not need to be stabilized by justification, in order for the regress to be brought happily to a halt. It would do for you to believe that, quite independently of whether you have any evidence for it, *qn* is a proposition which you cannot help believing. It could even be that for *every* proposition which you believe and believe to be evidence, you believe either that this proposition is one that you cannot help believing, or that there is evidence for it which you cannot help believing, or that there is evidence for evidence for it which you cannot help believing, or . . . , and so on. In this case, even if you do adduce, as evidence for a proposition you believe, a further proposition which you believe you *can* help believing and therefore want causally reinforcing justification for believing, the pursuit of this justification will sooner or later lead you to adduce as evidence a proposition which you believe you cannot help believing and which you will therefore not want causally reinforcing justification for believing. What more particularly might you believe, if you were in this fortunate position? It would do for example if you were a certain kind of extreme empiricist, if you believed that you cannot help believing propositions to the effect that you are having such and such an experience, and that the only propositions which are evidence for other propositions are either such experiential propositions or propositions which are supported by such experiential propositions, or propositions which are supported by propositions which are supported by such experiential propositions, or . . . , and so on.

There is the objection that if you now start reflecting about what makes you so fortunate, you will after all end up confronting yourself with an incompletable series of justificatory demands, albeit not the same series which you would have confronted yourself with if you had wanted to be justified in believing everything you adduced as evidence. For it is one thing to believe that you cannot help believing some proposition, another to believe that you cannot help believing that you cannot help believing it. If you believe that it is only because you believe that you cannot help believing some proposition *q* which you adduce as evidence, that you do not want the justification for believing *q* which would causally reinforce your belief that *q*, then you are liable to wonder about the stability of your belief that

(1) You cannot help believing *q*

Even if you also believe that

(2) You cannot help believing (1)

you are liable to start wondering about the stability of your belief that (2), and so on. There is a series of propositions (call it the *q*-series) whose first member is *q* and each subsequent member of which is formed by putting 'You cannot help believing' in front of its predecessor. However large an initial segment of the *q* series you believe, you are always liable to start wondering about the stability of your belief in the last member of that segment. Suppose that there is a member of the *q* series which you believe but do think you are able to avoid believing. Then you will want justification for believing that member. If you cannot find justification for believing it, then your overall position is not as satisfactory as it initially seemed. If on the other hand you can find justification for believing this member of the *q*-series, by finding a proposition *q*1, which you believe and believe to be evidence for it, then a question arises about the stability of your belief in *q*1, and we have the same problems all over again, this time with a *q*1 series.

But there is a proposition by believing which you could avoid confronting yourself with this new series of justificatory demands. Suppose you believe that

(H) All the beliefs you hold about which of your beliefs are unavoidable are themselves unavoidable.

You can then infer that every member of the *q*-series is a proposition which you cannot help believing. And if you wonder about the stability of your belief in (H) itself, you can reflect that the conjunction of (H) and the proposition that you believe (H) entails that (H) is itself a proposition which you cannot help believing.

III GLOBAL SUSPICION

Now consider finally the case in which, instead of starting out by wondering about the truth-value of some particular proposition, or about the future stability of the beliefs you now hold, you begin by standing back from and reflecting about the truth-value of your belief-system as a whole. You want that system to contain only true beliefs, and you want somehow actively to assure yourself that this is how it really is. Grant, as we initially supposed, that if you were to assure yourself of this, and were conscious of the process by which you had done so, you would recognize that you had *evidence* that the system contained only true beliefs. Grant also that if you had this evidence, you would *believe* the proposition q which was your evidence. Now either you already believe that q, at the time when you embark on your justificatory enterprise, or you do not. If you do, then an actual member of the whole set of propositions of whose truth you wanted to assure yourself would have to be evidence for that very set. And you cannot assure yourself that a conjunction is true by citing as evidence for its truth a proposition which is a member of that same conjunction. That would be to beg part of the very question at issue. If on the other hand you do not already believe the proposition you now cite as evidence, you do now believe it if you cite it as evidence, and there is a new set of beliefs, consisting of the old one plus this new belief, from which you can once again stand back, and of whose veracity you can once again want to assure yourself. The same dilemma is applicable to this new set, and so on ad infinitum. Now getting evidence for the new set requires you to get evidence for the proposition which was your evidence for the old set. So it seems that you are indeed liable to want a justified belief in the old set, of a kind which you cannot get without completing an infinite process of getting evidence for your evidence.

But now there is an objection. My treatment of the local uncertainty case depends on the principle that if you believe something you cannot simultaneously wonder whether it is true. And yet if that is right, it will be equally impossible for you to believe something and at the same time seriously want to assure yourself that it is true. How then can you want to assure yourself that your whole belief system contains only true beliefs, unless you somehow forget that you are yourself the owner of the belief-system about which you want to be thus assured? Global suspicion seems to require 'bad faith'.

The answer seems to be as follows. In the local uncertainty case, in which a proposition was adduced as or later thought to be evidence, and believed on that account, the belief was conscious, involved consciousness of or adversion to the particular proposition believed. And one cannot consciously believe something and simultaneously wonder whether it is true. But there is no reason why one should not now wonder about the truth-value of the whole system of non-conscious beliefs which one now has. One might well have any number of beliefs without being conscious of any of the propositions which someone else would say that one then believed.

This answer will generate a further objection. If you did assure yourself of the truth of your non-conscious belief system, what kind of belief, conscious or non-conscious, would you have in the proposition q which was your evidence? Presumably you would have a conscious belief that q. Now the regressive difficulty depended on this idea. If the belief that q already belonged to the belief system whose veracity it was to be evidence for, then it could not after all be such evidence. And if it did not already belong to this belief system, there would be a new system, consisting of the old one plus the belief that q, which would be just as much a focus of suspicion as the old one. And so on. Now we had to agree that any belief system of your own which you can stand back from and wonder about the truth of is a system of non-conscious beliefs. But the new system differs from the old system only by containing the belief that q. So if your belief that q is, as we are supposing, a conscious belief, what different and feasible object of suspicion can the new system present? I concede that if you got to the point of actually testing your belief-system, you would consciously believe any proposition which you then put forward as evidence, and could not then wonder whether that proposition was true, and construct a fresh object of suspicion by adding your belief in that proposition to the system you began by wondering about. But I think that so long as you do not *now* know what proposition it is that you *would* put forward as evidence and thus consciously believe, you can now wonder whether that proposition would be true, and so now wonder whether the new system, consisting of the old system plus the new and so far unidentified conscious belief, contains nothing but truth. You can still now be gloomily struck by the incompletability of the process on which you now want to embark, by the fact that whatever currently unidentified proposition you would cite as evidence in order to allay your suspicions of your whole belief-system, there is already just as much reason to want evidence for that proposition as there is to want to cite it as evidence in the first place. I conclude then that there is some melancholy interest in our initial sceptical train of thought.

School of Economic and Social Studies,
University of East Anglia,
Norwich NR4 7TJ

© N. M. L. NATHAN 1987

SOLIPSISM REGAINED

By MARK LEON

NEW approaches to old questions do not always escape the pitfalls associated with their predecessors. In part this is what I will attempt to show in examining one current, naturalized approach to epistemology.

First the traditional account. Some have sought in foundationalism a way of escaping solipsism. The thought is this: if some beliefs (foundational ones) could be shown to be certain, then other beliefs (higher-order ones) could be shown to constitute knowledge on the grounds of having some reasonable warrant. However, I rather tend to think of foundationalism as leading us into solipsism. Here is how it goes. Foundationalism requires – this is the only version I am considering – that before a belief can count as knowledge it must be incorrigible, infallible, certain, or such as could not turn out to be false. The primary case of beliefs which might satisfy this criterion are beliefs about oneself, more specifically beliefs about one's inner or psychological states. Or so it was generally thought. These, accordingly, would be the foundations. I am not sure that I go along with this; but that is no matter. What I do not go along with is the next step, the thought that if there are foundations, they enable knowledge of what lies beyond one's incorrigibly accessible ken. I think that even if there are foundations, there is no reduction of the rest of our supposed knowledge to them. Hence a form of solipsism. All that one can know about are one's own inner or psychological states. (Some might think that too quick. If there are foundations why must it follow that the rest of our beliefs have to satisfy the foundational criteria? Why can't there be inductive support for them? Foundationalism on this view would require that some beliefs be certain, not all. My response is that I accept the view that the rest of our beliefs have inductive, that is, fallible support. But it seems to me that if they count as knowledge we don't need foundations in the first place; and if we are committed to infallibility for there to be knowledge, then they don't count as knowledge. Either way it is foundationalism which fails.)

Traditional epistemology sought (philosophical) foundations for knowledge. Contemporary epistemology under the guise of epistemology naturalized lets foundations fall where they may within a naturalistic setting. The version that I am concerned with focuses on the notion of content and locates content within the world. It takes the content of a representational state to be a function of a certain causal condition which prompts it. This might seem an unlikely starting point for solipsism; still I shall argue there is a perfectly natural slide from these premises to the thesis that all we can have knowledge of is our own states. Naturalizing content is a feature of that thesis I take to be extrinsic realism; extrinsic realism takes content to be a function of the external or extrinsic conditions which (typically) generate one's representational states. Just as a thermometer represents heat because as a first approximation heat is causally responsible for the level of mercury, so our states gain representational content because of the causal conditions which prompt them. (For a more detailed account of extrinsic realism, see my 'Realism, Scepticism (and Empiricism)' forthcoming 1987).

Under what conditions do we ascribe a certain concept to some creature? Under what conditions do we take them to have a state with a certain representational content? Consider Dretske's account: '... possession of the concept *X* requires the possessor to have the resources for picking up, processing and responding to the information that something is *X*'. (Dretske 1983 p. 17. As will be seen I am indebted to Dretske's paper even though I am more sceptical than he of extrinsic accounts coming up with a satisfactory notion of, at least, perceptual content.) Dretske develops his account by relating information to what can be learned. 'A message (i.e. some event, stimulus or signal) carries information about *X* to the extent to which one could learn (come to know) something about *X* from the message'. (Dretske 1983 p. 10). I'm not sure that this is a worthwhile addition to the account. I suspect the important notion in which the possibility of such learning rests is that of differential sensitivity. To have the concept *X* is at least to be differentially sensitive to *X* things or the *X*-ness of things. There are two aspects here; the one is the notion of sensitivity reflected in a type of (typical) co-variation; the other is the 'differential' feature introduced to preclude – if possible – co-extensive properties. As Dretske indicates (1983 p. 8) the key to possessing the concept *X* is to be responsive to the *X*-ness of things (and not, say, the *Y*-ness which might be co-extensive with *X*-ness under certain condtions). In effect what makes it information about *X*-ness and not *Y*-ness is that one is differentially sensitive to *X*-ness and not *Y*-ness. There are conditions under which the effects of these two distinct properties could be separated. Were there no such conditions, actual or possible, then the appropriate concept to ascribe on this account would be that of *X* and *Y*.

Whether a state of some creature or device could misrepresent is a function of what it does represent. Typically we take some of our states to misrepresent – we think we have false beliefs, nonveridical experiences, and so on. But is this supposition in general correct? Is it possible for our states to misrepresent on this account of representational content according to which the content of a state is causally specified. The answer that I will suggest is that it is not possible.

Consider an altimeter. We take an altimeter to represent height above sea level because that is what it is sensitive to in the sense that changes in height are reflected (normally) in changes in the altimeter. Can it misrepresent height? At first blush it would seem that it could, for there could be a mismatch between the height indicated in terms of our calibration and the altitude it is at. But this is problematic. If we have already alighted on the height as that which is represented, this would be fine. But why not take – for this is a purely extrinsic account – the apparent mismatch to indicate that we have selected the wrong feature as that which the altimeter represents. Perhaps we have selected a property merely co-extensive with the property actually represented. Here is an analogy. We would ascribe, as Dretske suggests, the concept cylinder to a dolphin if it was able to pick out all cylindrical objects we put in its way. Assume all and only plastic cylinders were shown to it. Then in picking out the cylinders it also picks out plastic ones. So has it got the concept plastic? We think not. Not because it couldn't acquire such information about these objects, but because we think employing the Millean method of difference, we could separate out the property it is sensitive to in this case and the property co-extensive up to that point with the property it is sensitive to. We can indicate in this way that it was *cylindricality* rather than *plasticity* that the creature was sensitive to, and so exhibit that it was information about cylindricality which was encoded and so cylindricality which was represented.

What this indicates is that there are two things that go together; what we in fact are sensitive to, and what is not filtered out in employing the method of difference. What we are sensitive to is what we co-vary with under all conditions. What we do not co-vary with under all conditions might be inferentially signalled by association but is not what is represented. Effectively what we are sensitive to is what we are (nomically) infallible with respect to.

So return to the altimeter. Can it misrepresent height? It seemed so because under certain conditions it could indicate that the height is other than it is. For example, if the pressure were altered that would affect the reading we would get. Of course in normal conditions there is a correlation between change in the altimeter and the altitude it is at. But that should not be mistaken for its representing altitude – no more than being cylindrical should be mistaken for being plastic where the one is a reliable indicator of the other. (I don't mean to suggest that there is no difference between the two cases. Whereas in our world on the (unlikely) assumption that all and only cylindrical objects were plastic that would constitute a merely fortuitous connection between being cylindrical and being plastic, the relation between pressure and altitude is not – given the physical make up of our world – fortuitous. But this difference makes no difference to the point at issue here.)

Why altitude is not represented is because we can filter out the role of altitude. That suggests that changes in the altimeter are based in the first place on other changes. Information about altitude is inferred from other information. But following this method we eventually alight upon a property information about which is not dependent on information about some other property, is not, that is to say, inferential. And that is a property with respect to which the creature or device is infallible. There is no fooling the creature or device, for that possibility depends on going beyond the evidence depends on inference. But there is a base point for other inferences. And that is something with respect to which there is infallibility.

The point of infallibility is to rule out the extraneously though not fortuitously represented. If the altimeter's sensitivity to height is a function of its sensitivity to pressure, then it represents the latter not the former, though we can use it to measure the former. So what is represented is a function of what there is an invariant relation with respect to. Being infallible here is not some magical feature but expresses a brute nomic dependence. Relative to the laws of that world the one state cannot obtain without the other state obtaining, and the obtaining of the latter is sufficient for the former. Apparent misrepresentation is a function of using a state or device as an inferential signal for something else. Where there is inference, there is possible misrepresenation, but misrepresentation is something of a misnomer.

How does solipsism flow out of this? The process is all too easy. Take Dretske's example, the concept of a cylinder and apply it to us. The reason for thinking we have the concept of a cylinder and so for thinking that we have beliefs about cylinders is that we appear to be differentially responsive to cylinders. The presence of such items under appropriate conditions elicit a suitable response from us. Still as we saw with plasticity that might be a fortuitous coincidence. Is it possible for items other than cylinders to evoke the response? Of course it is. Non-cylindrical objects in abnormal conditions can look cylindrical; laser generated arrays of light in the absence of objects can elicit the same sort of experience; stimulation of various parts of our sensory system could do the same. So what does our experience represent? Not on this account the terminal point of the causal chain (taking it backwards, that is) but that point the occurrence of which is necessary and sufficient for the occurrence of the experience, that point which is not filtered out employing Mill's methods, or that with respect to which we are infallible. So what our experience encodes is not cylindricality, not even various structures of light, nor various of our sensory states. What it encodes, what is represents, is that state whatever it is which suffices for the occurrence of the experience. Such a state might indeed be a brain state. In this respect there is a departure from traditional solipsism; but such a state constitutes the boundaries of our knowledge because that is all that strictly speaking gets represented. We can at most know what we can represent. We can at most represent what we have infallible access to. At most we have infallible access to our own states. And so we have a certain form of solipsism. We can only have knowledge about (some of) our own states.

Creatures like us are sensitive to objects and properties in the world because of the changes wrought on the world and our perceptual apparatus by such objects. Because our perception of objects is in this uncontroverisal sense mediated, all the possibilities of error we normally associate with this process seem to flow in. But what those possibilities exhibit on this account is not actual error but misascription of content; failure to observe what in fact our states represent. Or such is the natural conclusion of one naturalistic account of representational content.[1]

University of the Witwatersrand,
1 Jan Smuts Avenue,
Johannesburg,
2001 South Africa

© MARK LEON 1987

REFERENCES

[1] Dretske, F. I., 1983. 'The Epistemology of Belief', *Synthese* 55.
[2] Leon, M. (Forthcoming 1987). 'Realism, Scepticism, (and Empiricism)', *Metaphilosophy.*

---

[1] I am grateful to the editor and an anonymous referee for helpful comments and suggestions.

TWO 'BY' WAYS

By MICHAEL WREEN

A FAMILIAR problem with the 'by' relation has been well put by Jennifer Hornsby:

> A man learns that certain particular muscles of his arm have to be contracted if ever he is to clench his fist; and we may imagine that he has a reason to contract those very muscles – perhaps he wants to please some experimenter. He does so. As we say: he contracts his muscles by clenching his fist.
>
> If we say that a man contracts his muscles by clenching his fist, do we not mean that his action of fist clenching caused his muscles to contract? He cannot bring about the contraction 'directly'; he had to bring it about by way of setting himself to do something else. But the reason why fist clenching serves as the kind of action he can carry out in order that he should have those muscles contract is that actions of fist clenching cause them to contract. So it seems. But attend for a moment to the mechanism

> of muscular contraction. Nerve impulses cause the contraction of muscles, which in turn cause the movements we make. We have it now not only that actions of fist clenching cause muscular contractions, but also that muscular contractions cause actions of fist clenching. Even if we do not wish to rule out causal loops *a priori*, it must come as a shock to learn that we encounter the phenomenon in this commonplace situation.[1]

Indeed, it would come as a shock to me, among many others, if this sort of situation did involve a causal loop. But I don't think we have to countenance such, at least not on the basis of the particular case under consideration here.

Hornsby manages to avoid causal loops and backward causation by invoking a distinction between transitive and intransitive uses of verbs like 'clench' and 'contract'. However, I shall argue that Hornsby's solution is mistaken and that the basis of the puzzle lies elsewhere.

The situation described is basically one in which we are tempted to subscribe to both:

(1) John clenched his fist by contracting his muscles.

and

(2) John contracted his muscles by clenching his fist.

But subscribing to both seems to entail a causal loop, or backward causation; hence, the case seems paradoxical. It seems paradoxical, however, only because (1) and (2) seem to entail backward causation; and I maintain they don't – that is, I maintain both are true but no causal loops or backward causation is entailed by their truth.

Rather than explaining this lack of entailment with a distinction between transitive and intransitive verbs, I shall argue that a different ambiguity is at work here, namely one in the 'by' locution itself.

In one use, the 'by' locution marks at least something like causation or what Alvin Goldman has called causal generation: John's contracting his muscles causes or causally generates his clenching his fist. In the circumstances envisioned, that is true, and its truth underwrites the truth of (1). Note, then, that both of the events described in (1) are actions: given the case description, 'contracting his muscles' has to be read as an act-type, properly prefaced with 'John's'. Contracting his muscles, after all, is what John was asked to do, what he set out to do, and what he did. The same is true of (2): both events described are actions, and the *same* two actions as described in (1); again, John both contracted his muscles (that is what he set out to do, and so on) and clenched his fist (he set out to do that, too, and did it). The point here is important, for if correct, Hornsby's solution to the puzzle won't do. For, according to her, the solution lies in distinguishing between transitive and intransitive uses of 'contract' and 'clench'; and according to her, 'actions can never be identified with bodily movements$_I$' (p. 5), that is, intransitive verbs do not designate actions. Thus, since only actions are involved in (1) and (2), and so in the generation of the puzzle, a distinction between transitive and intransitive verbs will not help resolve the paradox: there is no work for intransitive verbs to do.

---

[1] Jennifer Hornsby, *Actions*, Routledge & Kegan Paul, London, 1980, pp. 20-1. As Hornsby notes, much the same case is considered by Georg Henrik von Wright, *Explanation and Understanding*, Cornell University Press, Ithaca, NY, 1971, pp. 76-81, and Richard Taylor, *Action and Purpose*, Prentice-Hall, Inc., Englewood-Cliffs, NJ, 1966, pp. 194-95.

To return to the puzzle proper: (2), on the other hand, does not describe causation or causal generation at all. Understood in context, what it describes is something about John's action plan, and it records its successful completion. What John planned to do, the ultimate aim he sought to achieve, was to contract his muscles. But, as the complete case description Hornsby provides makes evident, he knew of only one way to achieve this aim, namely by taking direct, conscious, intentional aim at clenching his fist. Thus, he aimed at clenching his fist, knowing full well that because of his constitution, if he clenched his fist, that would indirectly achieve his having contracted his muscles. John's action plan, in short, operates in much the same way that an *a fortiori* proof plan does in mathematics: in order to prove that all natural numbers have a certain property, for example, Walt may, because he doesn't know how to proceed to that conclusion in any other way, set out to prove, and succeed in proving, that all real numbers have that property. Thus, to draw the analogy closer still, consider:

(3) Walt proved that all natural numbers have property P by proving that all real numbers have property P.

In the case of friend John, of course, the relationship between his two actions is not logical (i.e., there is no deductive relation between the sentences which describe his two actions, as there is, or would be, with a properly parsed (3)) but causal (i.e. there is an inductive relation between the sentences). But he, just like Walt in the hypothetical proof example just considered, deliberately overshoots, has an *a fortiori* action plan, which he uses to accomplish his main task. Thus, understood in context, what both (2) and (3) record is something about the indirect achievement or performance of an action.

That this way of interpreting (2) is correct can be reinforced, I think, if we consider what we would say about (2) if it were asserted without an accompanying story of the sort Hornsby provides. Asserted out of context, (2) would be considered to be plainly false. For in the absence of any indications to the contrary, (2) would be most naturally read with the 'by' marking the sort of causal relation, or relation of causal generation, that it does in (1); and so interpreted, the claim seems obviously incorrect, nay, bizarrely incorrect. (Commonsense but little tolerates backward causation.) In context, though, it is read as true, and thus whether the interpreter consciously knows as much, in much the same way that (3) is (or at least the way that (3) is by people who know a bit of mathematics, i.e. by people who can supply the relevant context, or background, easily enough); it is, in other words, like (3), read as a case of success through overshooting, or better still, indirect accomplishment or performance.

One word more, on knowledge and action plans. First, in the case of fist clenching, John knew that if he clenched his fist, he would contract his muscles. Interestingly, in this and many other cases – I don't say all – if the agent didn't have certain knowledge, we might be loath to attribute certain actions to him. If a non-culpably ignorant would-be pugilist hadn't been told about his physiology, then even if he spent a great deal of time clenching his fist, we might be reluctant to say he ever contracted his muscles, i.e. performed that action. The random scribblings of a small child, after all, do not a completeness proof make, regardless of their similarity to the scribblings of logicians. If, then, there is some hesitation in the case of our non-culpably ignorant would-be pugilist, the explanation probably has to do with (somewhat loose) conceptual connections between action and responsibility, and responsibility and non-culpable ignorance. Non-culpable lack of knowledge not only releases an agent from responsibility for an action but, in some cases, also releases him from performance of the action itself. (And similarly with intentions and beliefs: lacking an intention to deceive, or a belief that I can or will deceive, I cannot perform the action of lying or forging.) The small child proved nothing and, by parity of reasoning, some might argue – I do not want to take a stand on the matter – our ignorant pugilist didn't perform the action of contracting his muscles. A decision on the case of the ignorant pugilist is not needed here, however. For if John were our ignorant pugilist and the action of contracting his muscles should be attributed to him, the solution proferred above will do; but if it should not be so attributed, the paradox wouldn't arise, as both (1) and (2) would be false.

Second, I've emphasized 'success' and 'action plans' above, but the notion of an action plan is not, as such, central to the solution of the paradox. Obviously someone can, e.g., prove that all the natural numbers have a certain property by proving that all the reals do, even if his action plan didn't call for proving anything about all the naturals. Under certain conditions, much the same holds for (2); but those conditions include – importantly – being able to say that John performed the action of contracting his muscles. Thus, as in the last paragraph, if John lacked the action plan in question and, for whatever reason, the action of contracting his muscles is not attributed to him, the paradox doesn't arise; but if it is so attributed, then action plan or no, the solution proffered above will do. For central to that solution is the notion of indirect achievement, accomplishment, or performance, not action plans as such: the 'by' in (2) should be read as 'and this was indirectly achieved or performed by', the first disjunct perhaps being more appropriate when an action plan is present, the second when none is.

To sum up: causal loops and backward causation can be avoided, then, and the paradox dissolved, by noting that the 'by' locution is ambiguous in a way hitherto unnoticed. It can indicate not only an order of causal (or other sorts of) action generation – the 'by' in (1) means 'and this was caused (or causally generated) by' – but also something about indirect achievement or performance – the 'by' in (2) means 'and this was indirectly achieved or performed by'.[2]

Marquette University,
Milwaukee, Wisconsin 53233, U.S.A.

© MICHAEL WREEN 1987

---

[2] My thanks to the editor of *Analysis* and an anonymous referee for a number of useful suggestions on an earlier draft of this paper.

ANTIPHON

By GERALD VISION

THE indexical thesis, as originally named [5], holds that the sentence 'I am here now' will be true whenever asserted. In my first counterexample (henceforth (1)), I pointed to a mapped locatin other than my current one while asserting the sentence. In my second, (2), I left a note with this sentence downstairs, intending it to be read two hours hence when my wife arrives; but instead of remaining upstairs, I distractedly left the house without remembering to pick up the note. Professors Colterjohn and MacIntosh [2] believe these are sound objections to the thesis only when the referring terms in the sentence are used *demonstratively*, but not when they are used *indexically*. This contrast is the lynchpin of their reply, but they say very little toward explaining it and it is not, to my knowledge, an established philosophical dichotomy. My suspicion is that it is born of confusion, and I shall commence to explain why.

A question to be raised at the outset is 'Do we have one unambiguous sentence or two that happen to sound and look alike?' (Talk of *uses* rather than *senses* or *meanings* fudges just that issue.)

Evidence favouring the latter interpretation is their suggestion that English might have contained distinct words for each use. They imply that were this so it would not change the essential character of the situation now confronting us. Evidence that they believe these are distinct uses of a single sentence (-type) is their use of a single symbolization, S, for 'I am here now' and their customary talk of *the* sentence and *the* word 'here' (or 'now'). Moreover, whatever contrast they have in mind seems too local for a meaning difference. It certainly doesn't seem to conform to the differences present in paradigmatic cases of meaning equivocation (e.g. 'bank', 'bear'). Finally, I cannot conceive any grounds warranting, say, distinct senses of 'here', either in what the authors claim or generally. But since they give no grounds for supposing there are semantically distinct accounts of 'here', and the onus to do so would be on them, we needn't be concerned further with a view they may not wish to defend.

Let me illustrate why this issue matters. Earlier I claimed, as the authors note, that it is wrong to say that 'here' always refers to the vaguely bounded present location of the utterer. I was criticizing a semantic account of 'here'. The authors claim their distinction saves that account, for it is true at least of indexical uses of 'here'. If distinct *uses* don't lead to semantic differences, how could this be relevant? My claim concerned a certain doctrine about 'here', and didn't mention one or another of this single word's uses.

Let us then suppose that 'I am here now' is the same sentence for both putative uses. What would be an indexical, as distinct from a demonstrative, use of its expressions? Peirce, who introduced 'indexical' as a term of art, intended it, roughly, the way the authors take 'demonstrative'. As Arthur Burks [1] writes in a classical account of Peirce's notion, '[t]he symbol "this" is also an index, because ... it may function very much the same as the act of pointing; i.e. instead of pointing to a tree one may use the phrase "this tree".' David Kaplan [3], one of those I cited as maintaining the thesis, distinguishes indices from demonstrations, but only because the former are instances of a more general notion, embracing truth-value relevant contextual factors for terms other than explicit demonstratives. But on this notion a demonstrative can no more be contrasted with an indexical than can an ape with a mammal. However, if the authors are appealing to such model-theoretic considerations for their notion of an indexical use, other problems await. The procedure does not supply a novel way of linking a certain sort of term with coordinates, but assumes that issue settled (by whatever means) when attaching indexical structures to expressions. The decision cannot be based upon a distinction between indexical and demonstrative uses, for the relevant indexical use is no more than the result we wanted our canvass of all uses of 'here' to capture. If (1) establishes anything it is not something about one of several established uses of 'here', but about how a certain method of fixing indices fails to incorporate legitimate uses of the expression indexed.

These are only guesses about what the authors may have intended by their distinction, for they do not explain themselves. Moreover, their own account grows critically murky when discussing demonstrative uses. Where they cannot find a clear pointing, they conjure up surrogates that lack independent plausibility as ostensive devices. For example, they suggest two construals for (2) which would make it true without being a counterexample to the indexical thesis. On the second of these the piece of paper on which the sentence is written is selected as a 'proxy finger'. (Incidentally, they claim that *each* of their construals is a 'more plausible' interpretation than mine, including one on which I am regaling my wife with a philosophical truism that holds only because the 'here' and 'now' happen to be true of me wherever I am then. I can only remark that I am astounded by their intuitions about plausibility.) Their other example concerns a typist in Canada who teletypes 'I am here now' to be read in England *with the intention of deceiving the reader that the typist is in England.* They say that for this to be construed as a falsehood the typist must 'invoke a convention of ostension'. What convention and what ostension? Once again, I find the 'proxy finger' notepad and the typist's evil eye highly implausible candidates for ostensions. Is there anything to recommend them other than the authors' increasingly dubious thesis about the distinction between indexical and demonstrative uses? What explanatory weight does either carry?

The dispute can be rearranged as follows. Had I wanted to offer a post mortem, I might have claimed that holders of the indexical thesis concentrated only on cases in which 'here' and 'now' were used with minimal commitment, and in doing so had neglected instances such as (1) and (2) where this was not so. A one-sided diet of examples, a hallmark of over-hasty generalization, would be my diagnosis. Curiously, I think Colterjohn and MacIntosh would agree up to a point. Their counterclaim is that the two sets of cases, those on which holders of thesis concentrated and those which I brought to their attention, can be labelled 'indexical' and 'demonstrative' respectively. But their suggestion only saves the thesis if it can be shown that the classification has a clear and independent basis: that it is not *merely* a way of re-labelling the classes of non-counterexamples and counterexamples, respectively. I have been arguing that they haven't succeeded. For, first, admitting that S is one sentence, the authors must tread the very narrow and unlikely path of demonstrating a contrast sharp enough for their purposes, but not so sharp as to issue in a semantic distinction. Second, the authors tend to extend the notion of demonstration to cases where it seems arbitrary and certainly is not independently motivated. Finally, the terms as they are used in the philosophical traditions with which I am familiar fail to support any such contrast. I have no objection to the authors renaming the cases; but as a device to save the thesis it seems on a par with reclassifying black swans to save the generalization that all swans are white.

Mr Paul Simpson [4] also takes issue with my supposed refutation of the indexical thesis. I am unable to discover in most instances just what he is disagreeing with. Several remarks lead me to believe that he supposes the issue hangs on a difference between 'public' and 'private' displays of location. I do not know what he thinks of private displays, but for my purposes the utterer in (1) might just as well have uttered the sentence to himself with respect to a mislocation on a merely imagined map. For my part, I don't see that I have complicated an issue rather than clarified one by employing 'here' with an ostension. It is worth noting, just because of the semantic accounts of 'here' — and consequent misapprehensions about the way indexicals-demonstratives generally work — that deny that I can use 'here' sensibly while allowing it to mark some place other than my specious location.

Similarly, I do not see how I am providing 'now' with 'less mobility ... than dictionaries grant'. I can't even see the basis for the claim, since I make no attempt to give an account of 'now', provide only one example, and only say of that example that it is a natural, not an exclusive, interpretation of the utterance. In any event, Simpson's remarks about my use of (2) are unconvincing. He says my note is either a *prediction*, and thus *misinforms* because of its tense, or it is not, in which case it is true of me at the time of utterance. In all this Simpson has assumed that time of utterance (or writing, in this case) is time of statement, which is precisely what is being contested. That is why his one sample definition of 'now' ('at the time referred to') can be seen to fit my purposes as well as his. Ignoring that, though I have no objection to his extending the term 'prediction' however he wishes, not all statements about the future fall into the ordinary ambit of this category. (Is a will a prediction?) It would have been greater misinformation in the circumstances to have left the future-tensed message 'I shall be here'. If one looks without blinkers at the case, it is quite obvious there is nothing conventionally misinformational about the note (though it in fact misinforms in the case). Rather than using these evident facts to tell against his rigidly applied categories ('prediction', 'misinformation'), Simpson uses the categories unpersuasively to try to unsettle what still seems to be the only natural rendering of the note.

Evidently Simpson wishes to claim my counterexamples are mistaken, but other than the above attempt to overturn (2), which I have given reasons for rejecting, I can't find any grounds for his saying so. I don't see how the 'departed soul/time traveller' case bears on my critique, since I haven't implied anything inconsistent with 'here' and 'now' being public in Simpson's sense. And as for Simpson's last remark, that these words '... are detachable from utterances or acts of utterance and cannot be analysed completely, if at all, indexically', I should think that was a problem for holders of the indexical thesis if anyone. (I'm not sure it is a problem for anyone in this dispute, since the analysis of an expression needn't completely settle what it in fact refers to.)

A final word. I have the impression that behind both criticisms is the view that holders of the thesis could have crafted a more careful sentence that would have worked. 'Here$_1$' could have been defined as *wherever the utterer is at time of utterance*, and similar definitions would work for 'I' and 'now'. It was not part of my original task to prohibit this possibility. There may be many such self-guaranteeing utterances. (I didn't even comment on the several popularly ascribed to Descartes.) Rather, I set out to show that this is not so of a certain current English sentence, and that believing otherwise rests on myths about the ways of working of its terms, whether called indexicals or demonstratives. To protect the sentence from non-truthful utterances is a matter of stipulation, not a guarantee that is somehow naturally built into our forms of expression. I criticized holders of the thesis not for believing that their goal could be achieved, but for believing it was achievable with those tools. I'm not sure how significant a lesson this is, but I can vouch that the mistakes which all the critics seem to agree were uncovered (though they tend to minimize them) are errors of longstanding and broad circulation.

Temple University,
Philadelphia, Pennsylvania,
19122, U.S.A.

© GERALD VISION 1987

REFERENCES

[1] A. Burks, 'Icon, Index, and Symbol', *Philosophy and Phenomenological Research*, 9 (June 1949).
[2] J. Colterjohn and D. MacIntosh, 'Gerald Vision on Indexicals', *Analysis*, 47.1 (January 1987).
[3] D. Kaplan, 'On the Logic of Demonstratives', *Journal of Philosophical Logic*, 8 (1978), see esp. pp. 81-2.
[4] P. Simpson, 'Here and Now', *Analysis*, 47.1 (January 1987).
[5] G. Vision, 'I Am Here Now', *Analysis*, 45.4 (October 1985).

METAPHILOSOPHY

Edited by Terrell Ward Bynum

Metaphilosophy publishes articles and book reviews in a wide range of philosophical topics including the foundations, scope and function of philosophy, aspects of philosophical systems, applied philosophy, the sociology of philosophy, the teaching of philosophy, presuppositions of philosophical schools, and the relation of philosophy to other disciplines.

January, April, July, October
(two single issues and one double issue per year)
Subscription Rates Volume 17, 1986
Individuals: £19.80 (UK) £24.75 (overseas) US$41.80 (N America)
Institutions: £36 (UK) £45 (overseas) US$75 (N America)

Please send orders with payment to:

Iris Taylor
Journals Department
Basil Blackwell
108 Cowley Road
Oxford OX4 1JF
England

or to: Journals Department
Basil Blackwell
PO Box 1320
Murray Hill Station
NY 10156
USA

Basil Blackwell · Oxford and New York

Women and Philosophy
Special Issue

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Edited by Janna Thompson with the assistance of Mary McCloskey Brenda Judge San MacColl Elizabeth Gross

ARTICLES

Moira Gatens .... Rousseau and Wollstonecraft: Nature vs Reason
Gareth B. Matthews .... Gender and Essence in Aristotle
Karen Green .... Rawls, Women and the Priority of Liberty
Marion Tapper .... Can a Feminist be a Liberal?
Lorraine Code .... Simple Equality Is Not Enough
John Howes .... Mill on Women and Human Development
Susan Parsons .... Feminism and Moral Reasoning
Jan Crosthwaite and Christine Swanton ...... On the Nature of Sexual Harassment
Robert Pargetter and Elizabeth W. Prior ..... Against the Sexuality of Reason

CRITICAL REVIEW

Val Plumwood .... Ecofeminism: An Overview and Discussion of Positions and Arguments

PRICE: A$15. All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED BY THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

The Journal of Philosophy

SUBSCRIPTIONS

Individuals	$20.00
Libraries and institutions	$25.00
Students, retired/unemployed philosophers	$15.00
Postage outside U.S.	$ 5.00

BACK VOLUMES

Published bi-monthly 1904–1976, monthly thereafter.

All back volumes and individual issues are available, also a Cumulative Fifty-year Index, 1904–1953 and a Ten-year Supplemental Index, 1954–1963. Back volumes 1976–present also available in microform.

Please inquire for price lists and shipping charges on back orders.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol. 65, No. 1; March 1987

EDITOR: BRIAN ELLIS

ARTICLES

Frederick W. Kroon	Causal Descriptivism
Robert J. Lipkin	Altruism and Sympathy in Hume's Ethics
George Schlesinger	Accommodation and Prediction
Raymond D. Bradley	Wittgenstein's Tractatarian Essentialism
Jordan Howard Sobel	Self-Doubts and Dutch Strategies
Raymond Martin	Memory, Connecting and What Matters in Survival
Gerald H. Paske	Sperm-Napping and the Right Not to Have a Child

DISCUSSION

B.J. Copeland	Appraising Historical Accounts: A Discussion of Gorman's View

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTION: Individuals A$20 (US$20 or Stg 10); Institutions A$40, (US$40 or Stg 20). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

NOTES

The Analysis Committee consists of: Chairman, P. T. Geach; Secretary, T. R. Baldwin; T. S. Champlin, M. Clark, J. P. Dancy, Andrew Harrison, B. J. Heal, R. F. Holland, D. H. Mellor, H. W. Noonan, P. Smith, M. A. Stewart, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

Subscriptions. Analysis is published by Basil Blackwell Ltd, 108 Cowley Road, Oxford OX4 1JF, England and P.O. Box 1320, Murray Hill Station, NY 10156, U.S.A. Each volume comprises four numbers, each of at least 48 pages, appearing in January, March, June and October. Orders should be sent to the publisher. Subscriptions prices for 1986 are as follows:

	U.K.	Overseas (incl. Eire)	U.S.A. & Canada
Institutions	£12.00	£15.00	U.S. $33.00
Individuals	£8.00	£15.00	U.S. $21.00

Back Issues. Volumes 1–37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

Microfilm copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30–32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

Advertising. For details contact Bill Drake, telephone (0869) 38477, or write c/o the publishers.

Contributions. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributions are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable

Vol. 47 No. 3 (New Series No. 215) June 1987

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

Determinism, predictability and chaos	G. M. K. HUNT
Moral realism and the modal argument	ROBERT ELLIOT
Williams, *ought*, and logical form	RICHARD LEE
Assumptions and the supposed counterexamples to modus ponens	D. E. OVER
Free and classical counterparts: response to Lewis	GRAEME FORBES
Lewis on perdurance versus endurance	E. J. LOWE
On knowledge and the unknowable	TIMOTHY WILLIAMSON
Salmon versus Kripke on the *a priori*	RALPH KENNEDY
On Lemke's defence of a causal basing requirement	JONATHAN L. KVANVIG
A first-person asymmetry	JAMES MONTMARQUET
Sense, reference, and meaning-incommensurability	STIG ALSTRUP RASMUSSEN
Frege on infinite axiom-systems	R. R. ROCKINGHAM GILL
A puzzle concerning divestiture	STEVEN M. CAHN

BASIL BLACKWELL · OXFORD AND NEW YORK

ISSN 0003-2638

ANALYSIS 47.3 JUNE 1987

DETERMINISM, PREDICTABILITY AND CHAOS

By G. M. K. HUNT

THE thesis of physical determinism is often supported by an appeal to a mechanical view of the world. This mechanical view derives much of its support from an appeal to classical mechanics. In this paper I wish to discuss the recently discovered *chaotic systems* and show how their existence severs the link between classical mechanics and determinism. I shall first explain the nature of chaotic systems then go on to discuss their importance for the analysis of physical determinism.

A physical system can be described in many ways. In this paper I will use a *phase space* description. A phase space description represents a system at a particular time by assigning it to a point in a conceptual space. The coordinates of this space are the physical variables used to describe that system. In the case of a pendulum, for example, the variables are position and momentum of the bob. As the pendulum oscillates the variables change their values systematically and trace an elliptical path in the phase space (see Fig. 1 Path A).

If the pendulum had been launched from the central position with a slightly greater initial velocity than before, the path in phase space would begin close to the first path and stay close to

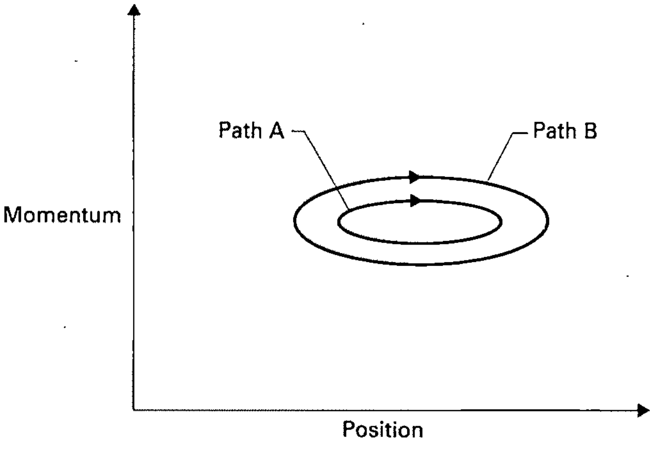

Figure 1

it throughout the ensuing oscillations (Fig. 1 Path B). Many physical systems, when launched with slightly differing initial conditions, have paths in phase space which remain close.

Another system, for example a ball rolled across a flat surface, but on different occasions in slightly different directions, will follow different paths whose separations increase with distance along the paths. Their paths in phase space will do likewise. But we can make their divergence at a later time as small as we like by making the initial difference small enough. Such systems are said to exhibit continuity in their behaviour under changes of initial conditions. By calling the behaviour continuous we draw attention to the fact that variation in their subsequent behaviour changes smoothly with variation in their initial state. It should be understood that we are commenting on the relation between the behaviour of pairs or groups of similar systems, or the same system on different occasions, rather than any one.

Not all systems exhibit this continuity. If a marble is balanced on top of a larger ball its path in phase space is simply a point; it maintains its position and momentum. If it were placed slightly to the left it would roll left and perhaps continue for some distance. A slight change in the initial condition has resulted in a movement to the left whose size cannot be reduced by reducing the size of the initial disturbance. But such points of bifurcation in the phase space have not troubled determinists. Any indeterminacy of prediction of the subsequent behaviour of the marble can be attributed to inaccuracy in the setting of the initial position. And for any particular placement of the marble there will be a sufficient accuracy of measurement of the initial position which will allow prediction of the subsequent behaviour. This will be analysed further when we discuss determinism and predictability.

The paths of these systems in phase space diverge according as they originate from the left or right of some critical dividing line. All the systems on the left, represented by an area of initial states, go in one general direction, all those on the right in another (Fig. 2a). Within the sub-areas continuity still applies: positions at a later time can be made as close as desired by making the earlier positions as close as necessary. This can only be so if a point which is between any two points in the initial area of phase space remains between those points in a later area. (The notion of 'between' used here is a complicated and topological, rather than a simple metrical, sense.)

Recently a class of systems have been discovered whose behaviour is much more complex and fails the continuity condition in a radical way. These are the chaotic systems. No matter how close two systems are initially, their phase space paths may diverge arbitrarily far. This can occur not just for isolated points like the marble placed centrally on the ball, but uniformly over a range of initial states. These sets of initial states may cover a substantial

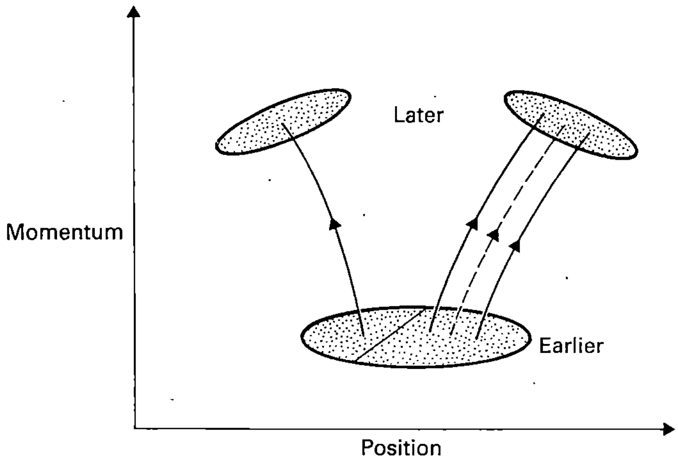

Figure 2a

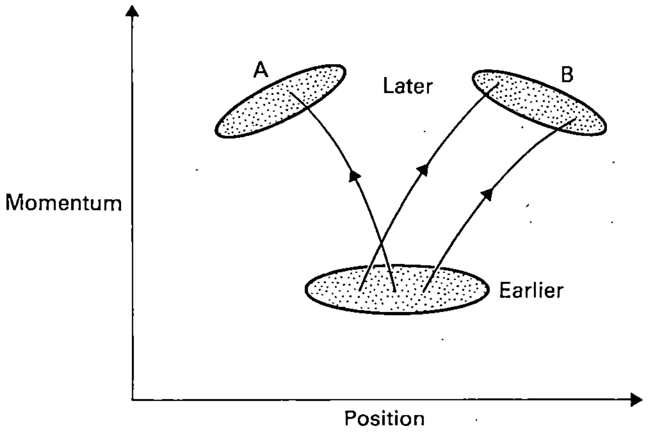

Figure 2b

area in phase space. Between any two points whose path ends later in, say, area A there will be a point whose path ends in B (Fig. 2b), and *vice versa*. This phenomenon will occur no matter how close the chosen points are. There is a dense intermingling of points whose subsequent paths become arbitrarily distant.

Even if the areas A and B were adjacent we still could not smoothly move the final position of a system by smoothly moving the initial position. The final position would jump about in an uncontrolled way. There is no continuity between changes in initial position and changes in final position.

To appreciate the effect of this dynamical behaviour on the thesis of determinism we must look closely at the relationship between physical and epistemic determinism.

Physical determinism is the thesis that the state of a physical system, perhaps the whole universe, is a definite function of its state at a preceding moment. This definition can be put epistemically by saying that if the state of a system at a particular time is known then its state at a later time can be predicted. This prediction can be made using the deterministic laws of classical mechanics. The success of classical mechanics as a predictive theory is therefore evidence that the universe is physically deterministic.

But epistemic determinism cannot be known to be perfectly true. We cannot know with complete precision that predictions have been precisely realized. This is widely known and accepted; the reasons need not concern us here.

Given the impossibility of perfectly accurate measurement we can restate the thesis of epistemic determinism in the following form. Predictions can be made arbitrarily accurate by making the determination of the initial state of the system arbitrarily accurate. In this form epistemic determinism still entails physical determinism. Until recently, this form of epistemic determinism had been found true and been used to support physical determinism.

But, as we have seen from our discussion, such a thesis requires continuity across sets of initial and final states of a system. And we have seen that chaotic systems do not exhibit such a continuity. If classical mechanics is true then chaotic systems are to be expected. If this is so, and accurate measurement is impossible, then epistemic determinism does not follow from physical determinism and is in fact false.

Classical mechanics remains a physically deterministic theory. Every single initial point in phase space is on a path leading to a single later point. But when we admit the impossibility of perfectly accurate measurement, epistemic determinism must depend upon the existence of a further relation — continuity — between the paths of single system started under slightly different initial conditions. It is the assumption of this further relation that is incompatible with the existence of chaotic systems.

The failure of epistemic determinism does not therefore undermine the thesis of physical determinism. The failure is, on the contrary, to be expected. In a world where perfectly accurate measurement is impossible, classical physical determinism does not entail epistemic determinism, even in the weakened form suggested.

University of Warwick,
Coventry CV4 7AL

© G. M. K. Hunt 1987

REFERENCES

Feigenbaum, M. J. Universal Behaviour in Nonlinear Systems, in *Physica* 7D, 1983, pp. 16–39.
Lasota, A. & Mackay, M. C. *Probabilistic Properties of Deterministic Systems*, Cambridge University Press, 1985.
Prigogine, I. & Stenger, I. *Order out of Chaos*, Fontana, 1984.

MORAL REALISM AND THE MODAL ARGUMENT

By ROBERT ELLIOT

IN several places Simon Blackburn has developed what might be thought of as the modal argument against moral realism. I shall discuss this argument in its most recent, and sophisticated, form[1] and show that it does not work in the way that Blackburn thinks it does.

Consider the following:

(S) $N((\exists x)(Fx \,\&\, G^*x \,\&\, (G^*x \cup Fx)) \supset (y)(G^*y \supset Fy))$

(?) $N((\exists x)(Fx \,\&\, G^*x \,\&\, (G^*x \cup Fx)) \supset N(y)(G^*y \supset Fy))$

(P) $P(\exists x)(G^*x \,\&\, {\sim}Fx)$

and assume

(E) $(\exists x)(Fx \,\&\, G^*x \,\&\, (G^*x \cup Fx))$

'*F*' is a term from the moral vocabulary. '*G**' is a term, or set of terms, from the natural or physical vocabulary. '$\cup$' indicates that the natural vocabulary underlies the moral vocabulary: it indicates some kind of supervenience by the moral upon the natural. 'N' is a modal operator expressing some kind of necessity and 'P' is a modal operator expressing some kind of possibility. In what follows two kinds of necessity and possibility are discussed; analytic (or conceptual) and metaphysical. So, for example, (S) may be offered as a claim about analytic necessity:

(S_a) $AN((\exists x)(Fx \,\&\, G^*x \,\&\, (G^*x \cup Fx)) \supset (y)(G^*y \supset Fy))$

or as a claim about metaphysical necessity:

(S_m) $MN((\exists x)(Fx \,\&\, G^*x \,\&\, (G^*x \cup Fx)) \supset (y)(G^*y \supset Fy))$

---

[1] See 'Moral Realism', in *Morality and Moral Reasoning*, ed. J. Casey, London: Methuen, 1971; *Spreading the Word*, Oxford: Clarendon, 1984; 'Supervenience Revisited', in Ian Hacking (ed.) *Exercises in Analysis*, Cambridge: Cambridge University Press, 1985. All page references are to this last work.

Blackburn says:

> If we put (?) into abeyance we should be left with a possible form of doctrine which accepts both (S) and (P): the (S)/(P) combination. It is this which I originally claim to make a mystery for realism ... there are relations between F and G vocabularies which are properly characterized by (S) and (P). In the moral case I think that this is best explained by a projective theory of the F predicates. (pp. 52–3)

The modal argument against moral realism depends on finding the (S)/(P) combination in the moral case. If (?) can be shown to be plausible in the moral context, then the moral realist is provided with a counter to the modal argument. I think that (?) can be shown to be plausible in this context.

First, something needs to be said about what is mysterious in the (S)/(P) combination. (S) imposes a ban on mixed worlds, that is on varying associations of *F* and *G* terms within particular worlds. In the moral case (S) prohibits the moral floating free of the natural. However, (P) entails that there will be some possible worlds in which there are different combinations of the natural and the moral from those in other worlds. In other words, there is a ban on mixtures within worlds but no ban on mixtures across the structure of possible worlds. The mystery concerns the authority behind the ban, as Blackburn puts it, 'on inter-world travel by things which are, individually, at home' (p. 53). The (S)/(P) combination seems 'to offend against a principle of plenitude with respect to possibilities' (p. 53).

Now Blackburn is exactly correct in his account of what is *prima facie* puzzling about (S)/(P) combinations. And if there is a genuine (S)/(P) combination in the moral case his modal argument would seem to favour anti-realism about moral properties. Perhaps the moral realist could develop an explanation of the puzzle compatible with his or her preferred view but things would be neater if the combination could be ruled out. And showing the plausibility of (?) would rule it out.

It is worth highlighting why the puzzle would disappear with the re-emergence of (?). Its effect would be to impose a ban on varying mixtures across the whole structure of possible worlds and the ban on mixtures within a world would simply be a particular instance of this more general ban. The apparent arbitrariness of the division of possible worlds required by (S)/(P) is not present if (?) applies in the moral case. (?) treats particular worlds and the whole structure of possible worlds in a singular fashion. While it may be that some puzzle remains about why there is this singular, systematic ban on mixtures, it is a different puzzle from the one which mediates Blackburn's argument against moral realism.

Let us now focus more precisely on the moral case and ask what support might be offered for (S) and (P). (S) is supported by the conceptual consideration that it is impossible to suppose that if two things share their natural properties in a given world they could nevertheless differ with respect to their moral properties. This is alleged to be an analytic or conceptual truth entailed by central features of our moral practices and discourse. Now this would seem to support (S) read as (S_a) (see p. 56). And (P), read as (P_a), is said to be supported by the consideration that the content of morality is not strictly determined by conceptual or analytic necessity. Blackburn says:

> it is not plausible to maintain that the adoption of some particular stand is 'constitutive of competence' as a moralist. People can moralize in obedience to the conceptual constraints that govern all moralizing, although they adopt different standards, and come to different verdicts in the light of a complete set of natural facts. (p. 56)

My argument against Blackburn has the following form. First, I shall argue that the reasoning which leads to the ban on mixed worlds leads to a ban on mixtures across the whole structure of possible worlds. In other words, if (S) is true for the reasons Blackburn offers then (?) is true. Second, I shall urge that the reasoning which is given in support of (P) can be extended to support ~(S). Either way, the (S)/(P) combination is eliminated in the moral case. Moreover, these arguments against the modal argument succeed even where the modalities are taken as analytic and not simply metaphysical.

Consider the reasoning which is said to yield (S_a). The claim is that there is an analytic or conceptual constraint on moralizing which compels us to treat cases where the natural properties are identical in morally identical ways. To do otherwise would be to engage in a practice other than moralizing. This is an explanation of supervenience, an explanation of (S_a), which is anti-realist or leastways fits well with anti-realism. Now this same analytic or conceptual constraint would seem equally to generate a ban on mixtures across the structure of possible worlds.

Imagine that I am asked whether I think that an actual state of affairs, S_w, is good. I notice that S_w is characterized by the set of natural properties G^* and I judge that it is good. I am then asked to imagine a possibility which is not actual. This possibility contains a state of affairs, S_p, which is likewise characterized by G^*. I am asked whether S_p is good and I judge that it is. Whatever moves me to judge S_w as good in virtue of exemplifying G^* should move me to judge S_p as good since it similarly exemplifies G^*. Moreover, the force of 'should' here is conceptual. States of affairs characterized by G^* are not found only in the actual world; they are scattered across the whole structure of possible worlds. Judging that S_w is good but not that S_p is good would just as much breach the conceptual constraint on moralizing as would failure to abide by the ban on mixtures within the actual world. It is 'constitutive of competence' with the moral vocabulary that mixtures be eschewed across the structure of possible worlds. The only other explanation of failing to adhere to this constraint is that the property of being located within some particular world, such as the actual world, plays the role of a releasing property, that is a property which is in fact included in the set of properties underpinning the judgement that S_w is good. This has analogues in instances where the mixed world ban is apparently breached. Here spatial or temporal location might function as releasing properties. However, these special instances do not subvert the overall pattern of argument which extends the conceptual constraint from the ban within worlds to the ban across worlds. And the latter, together with our assumption (E), entails $\sim(P_a)$.

It is important to emphasize that, in the example just now discussed, my evaluations of S_w and S_p are both made from my point of view in the actual world. Assuming that individuals are not world bound, there will be possible worlds in which I view S_w as bad. And from my point of view in such worlds I would be conceptually constrained to make the same judgement about S_p. From my point of view in the actual world there is a ban on mixtures across possible worlds. From my point of view in each possible world there is a ban on mixtures across possible worlds. However, this is no more difficult to appreciate than the fact that in the actual world my evaluation of S_w might differ from another's. The moral realist may say that this is possible because moral beliefs can be false. My argument so far is that the ban on mixed worlds is analytically necessitated because the ban across worlds is analytically necessitated. That is the situation to which the logical extension of Blackburn's argument for (S_a) takes us.

Interestingly Blackburn comes very close to making this same point. He says:

> ... if I hold that the fact that someone enjoys the misery of others underlies the judgement that he is evil, I should hold that in any possible world, the fact that someone is like this is enough to make them evil. (p. 55)

However, Blackburn thinks that what this rules out is (P_m) not (P_a). He thinks that it invites $(?_m)$ not $(?_a)$. If Blackburn were right about this the moral realist would be faced with the problem of explaining an $(S_a)/(P_a)$ combination. My claim is that the reasoning Blackburn offers in support of (S_a) is sufficient to support $\sim(P_a)$ and $(?_a)$. And $\sim(P_m)$ and $(?_m)$ have to be true because they are entailed by $\sim(P_a)$ and $(?_a)$.

My argument against Blackburn might be given a slightly different twist. Consider again his reason for supporting (P_a), namely that analytic or conceptual constraints do not compel the adoption of particular moral standards. This presumably is what leads Blackburn to permit mixtures across the structure of possible worlds. However, the very same line of reasoning would seem to permit mixtures within worlds. In fact the reasoning as Blackburn expresses it, 'people can moralize in obedience to the conceptual constraints that govern all moralizing, although they adopt different standards'

(56), says just this. This is just to say that diverse moral points of view can exist, not least because they probably do exist, within a particular world. So the reasoning that supports (P_a) seems also to defeat (S_a). The obvious line of retreat is to shift to the claim that (S_a) constrains judgements made from a particular moral point of view, for instance the point of view I have in the actual world right now. However, this leads nowhere since (P_a) has, in these circumstances, been closed off. My argument here is that to secure (P_a), (S_a) has to be dropped and that would remove the mystery the moral realist was thought to face.

I have, in effect, argued that $((S_a) \& \sim(P_a)) \vee ((P_a) \& \sim(S_a))$ in the moral case. While I think the modal argument for anti-realism fails, I do not think that moral realism is true. But that is another story which I don't quite know how to tell.[2]

Brisbane College of Advanced Education,
130 Victoria Park Road, Kelvin Grove,
Queensland 4059, Australia

© ROBERT ELLIOT 1987

---

[2] I would like to thank Simon Blackburn and an *Analysis* referee for their helpful comments and suggestions.

WILLIAMS, *OUGHT*, AND LOGICAL FORM

By RICHARD LEE

IN '*Ought* and Moral Obligation' (*Moral Luck*, Cambridge, 1981) Bernard Williams takes Gilbert Harman to task for suggesting a certain distinction among types of *ought* statements. In particular, Williams is critical of the suggestion that there is some difference in logical form between statements such as 'It ought to be the case that Johnson die for the murder' and ones like 'The district attorney ought to prosecute Johnson.'

Before exploring Harman's proposal, let us consider first, as a plausibly analogous case, the word 'clearly'. This word can appear in statements as a modifier of a clause, as in: 'Clearly, life doesn't go on forever.' But it can also appear as a modifier of a verb as in 'The police officer spoke clearly.' Notice that this latter sentence is different in meaning from 'Clearly, the police officer spoke.' The sentence with the adverb in front might be true, for instance, where the other sentence is not (if it was obvious that the police officer said *something*, but even those nearby could not make out *what* she said). And there might be sentences which are ambiguous between these two uses. Consider: 'The accused had clearly been told what his rights in the case were.' Thus the distinction between these two senses of 'clearly' may not show up as a distinction in the surface grammar of the sentences in which they occur. The difference, we want to say, is in the logical form of the statements in which these uses appear.

Similarly, it has been suggested, the term *ought* has two different logical functions. In addition to functioning as a modal verb or relational predicate, it can, apparently, act as a 'propositional operator'. This latter kind of function is not particularly mysterious. There are certainly other terms that are to be understood as propositional operators. When philosophers wish to emphasize that terms such as 'not' or 'necessarily' apply to complete clauses instead of to terms within those clauses, they concoct phrases such as 'It is not the case that' and 'it is necessarily the case that'. (One might want to emphasize, for instance, that one is not appealing to modal properties of objects by saying 'Necessarily, mathematicians are rational' instead of 'Mathematicians are necessarily rational.') To emphasize a propositional use of the term *ought*, then, *ought* judgments could naturally be expressed in a similar style: 'It ought to be the case that the treaty remain in effect until another one can be signed.'

Ought appears also to function at times as a modal verb or relational predicate, much like the word 'can'. When I say 'Vola can type ninety words a minute', it seems that I am asserting something to be true of Vola, that 'can type ninety words a minute' is being predicated of Vola, or, alternatively, that a certain relation is being asserted to hold between Vola and the action type 'typing ninety words a minute'. And when I say 'You ought to tell me what happened', it equally appears that I am predicating 'ought to tell me what happened' of you or asserting a certain relation between you and the action type 'telling me what happened'.

The position Williams is arguing against was presented by Harman in his review of Roger Wertheimer's *The Significance of Sense* (*The Philosophical Review* 82, 1973, pp. 235-9; essentially the same distinctions appear in Harman's article 'Reasons' in *Critica* 7, No. 21 (1975), reprinted with slight editing as chapter 10 of Harman's *The Nature of Morality*, Oxford, 1977; see also p. 85 of that book). According to Williams, 'Harman claims to detect two kinds of *ought* sentence, distinguished by a difference of logical form' (p. 114). Thus on Harman's view there are two differing senses of *ought*, one representing 'a property of a state of affairs', and the other representing 'a relation between an agent and a possible course of action' (p. 115).

Williams admits that the first, the propositional sense, must be allowed:

> It is incontestable that there is a use of *ought* in which it does not express a property of an agent. It occurs, for example, in

> (3) This room ought to be swept,
>
> which can be represented as of the form $O(p)$, with *ought* occurring as an operator on a proposition (p. 115).

But according to Harman there is also *another* sense of *ought* which cannot suitably be represented as $O(p)$. It is this claim that Williams is challenging.

Williams argues, *pace* Harman, 'that there is no reason to regard the *ought* of moral obligation as anything but a propositional operator' (p. 115). It is the soundness of Williams' argument to this point which I wish to call into question.

Williams sets it up like this:

> Consider the pair
>
> (5) Someone ought to help that old lady
>
> and
>
> (6) Jones ought to help that old lady.
>
> A claim such as (6) may, very familiarly, be supported by the claim (5), together with some consideration which specially selects Jones; for instance, that he is the only person within striking distance competent to give the required help (p. 116).

This is not yet something that Harman would need to disagree with. It does not yet undermine the distinction between the two logical forms. After all, Williams is allowing for the sake of argument that (5) contains a propositional operator use of *ought* while (6) contains a use of *ought* instead as a relational predicate.

But what Williams is trying to do here becomes clear in the next sentence:

> The occurrence of *ought* in (5) is a propositional operator, and it is hard to see what requires it, or even allows it, to turn into something else in (6) (p. 116).

What he is suggesting in this ironic way, of course, is that *nothing* allows it to turn into 'something else', that the *ought* in (6) cannot be other than it is in (5), where it is allowed to be a propositional operator. So the conclusion of Williams' argument here is that there are *not* these two different 'logical forms' for *ought* statements, and, more specifically, that (5) and (6) do not involve different senses of the word *ought*, as Harman presumably would maintain.

Williams' point seems to be a simple, logical one, namely that if a term appears in a premise and in the conclusion of an argument, then unless the argument is fallacious the term must not be used with a different meaning in the conclusion from the one it has in the premise. In the present case, if the *ought* which appears in (5) is a propositional operator and the *ought* which appears in (6) is a relational predicate (and not a propositional operator), then the inference would be fallacious. But it is not. Therefore the uses of *ought* in (5) and (6) must not differ in this way.

Let us see whether Williams is right about this. First we must look at his premises.

(i) (5) involves a propositional operator.

The argument that (5) involves a propositional operator must be that it cannot be equivalent to

$(\exists x)\,O(x$ ought to help that old lady),

since it might be true without there being such an x. That is, it might be true that while it is the case that *somebody* ought to help the woman, there is no one of whom we can correctly say that *he* (or *she*) should. That is to say, we cannot pin the responsibility on any one person (or on any group of people distributively), and this is not merely a limitation of our abilities: (5) may be true without there being anyone with such a responsibility. This all seems to be correct, provided that we allow that indeed (5) says something that can be true in these circumstances. After all, someone might doubt that any sentence like (5) would be *true* if there were no individual who ought to help the woman, thus questioning (i).

Of course this attack may in the end amount to denying *any* propositional operator use of *ought*, claiming instead that the only use of *ought* is to pick out a relational property. This, of course, would be denying what Williams claimed to be 'incontestable'. Such a critic of (i) would likely say that (3) is either meaningless nonsense or is equivalent to 'There is some person who ought to sweep this room,' which, admittedly, was probably not what was intended by the person who uttered (3). This line of attack on Williams differs from his position at this very first premise. But it is important to note that it is in no greater harmony with Harman's position, since Harman would surely accept (i), or at least accept that there are occurrences of the use of *ought* as a propositional operator. Indeed the critic here stands with *Williams* on the main issue: there is only one logical form of ought statements (at least in regard to this matter), not the diversity that Harman sees.

Let us, however, grant Williams this premise (particularly since his intended opponent must) and plough ahead. Consider the next premise, namely:

(ii) (6) is not a propositional operator.

This premise is used to set up the *reductio* argument, so we should grant it for the purpose of argument. Clearly Harman would accept (ii), and Williams is merely accepting it here to demonstrate a problem with Harman's account.

Consider then Williams' next claim:

(iii) From (5) and something one can derive (6).

As I've expressed it here, of course, this is trivially true. For any p and q, one can derive q from p together with *something*. But

Williams is more specific in what he thinks this 'something' is. Of course one can go from (6) to (5), but this would then yield an obviously relational sense of *ought* in (5) – and that is not the direction the deduction is supposed to move here. Now in spite of the fact that Williams characterizes the inference as 'very familiar', it is far from clear precisely what premise one could add to (5) to yield (6) – at least along the lines that Williams has in mind. But for the moment let us presume that this can be done. Still, I want to say, even granting all these premises, Williams' conclusion does not follow.

I challenge Williams' claim that the *ought* here is supposed 'to turn into something else.' Yes, the hypothesis (which Williams is attributing to Harman) is that the *ought*'s are different. And it is claimed that (6) follows from (5) with an additional premise. Fine, but nothing has been offered to suggest that somehow the *ought* in (5) *is* the one in (6) – or that it 'turns into' it. To see the point here consider how (6) would follow from (5). It would *not* follow with the mere addition of a premise that Jones is competent to help the woman and is within striking distance (although this is what Williams seems to imply). Rather there must be a premise like this:

> If someone ought to do *X* and *P* is the only person capable of doing *X*, *P* ought to do *X*.

Notice that there are *two ought*'s in this additional premise and there is nothing which prevents the first one from being an evaluative ought (a propositional operator in Williams' sense) while the second one is relational, involving obligation. Indeed, I suggest that (barring objections put aside under discussion of premise (i)) this is how the claims would naturally be understood. But if *this* is part of the 'something', the additional premise that needs to be added to (5) to get to (6), it is quite clear how the *ought* in (6) might be a relational term even while that in (5) is a propositional operator. The *ought* in (5) did not turn into 'something else' in (6) because *that* occurrence of *ought* does not reappear in (6). The *ought* in (6) comes from elsewhere – namely from a tacit connecting premise.

Once this is seen, it is obvious that the general structure of Williams' inference here is unsound. Just because a particular word appears in a premise and also in the conclusion of a valid argument, particularly an enthymematic one, is no reason to suppose the word has the same function in each. Someone might claim that 'It is right for Jones to help the old lady' together with some (unspecified) additional premise entails that Jones has a right to help the old lady. But we cannot conclude from this that the word 'right' has the same meaning in premise and conclusion. Of course if the argument is formally valid, then there must be in the premises *some* use of the word 'right' *in the same sense* in which it is used in the conclusion. But this requirement is easily satisfied if the suppressed premise connects the two uses of the word, as 'One has a right to do what it is right for one to do' does. And it is precisely this kind of connecting premise that appears to be suppressed in Williams' derivation of (6) from (5). Surely without such a premise, as the one I have supplied, the move is unconvincing.

Now Williams may try to claim that such a premise is not at all necessary to the inference. After all, we are happy saying that 'if not *q*, then not *p*' follows from 'if *p*, then *q*', without requiring the additional premise 'IF if *p* then *q*, THEN if not *q*, then not *p*.' But the cases are, of course, not analogous. The 'missing' premise in the contraposition example is logically true. Its inclusion in the argument is thus superfluous. Whatever connects (5) (together with the facts about Jones' ability and availability) with (6) is surely not a logical truth. Yet for (5) together with relevant facts about Jones to 'support' (6), *some* connection is required. I do not, therefore, think that claiming the inference was not intended as deductive would obviate the need for some connecting premise (or rule of inference, if you prefer it that way) of the sort I have indicated.

The upshot of all this is that while Harman may in the end be mistaken in his analysis of *ought* statements (since (i), at least, requires further argument), Williams has not settled the issue by showing him to be wrong in distinguishing two logical forms *ought* statement may take.

University of Arkansas,
127 Kimpel Hall,
Fayetteville, Arkansas 72701, U.S.A.

© RICHARD LEE 1987

ASSUMPTIONS AND THE SUPPOSED COUNTEREXAMPLES TO MODUS PONENS

By D. E. OVER

IS modus ponens a valid form of inference for the ordinary conditional? In a recent article Vann McGee has surprisingly argued 'no' on the basis of some very interesting examples. (See 'A Counterexample to Modus Ponens' in *The Journal of Philosophy*, September 1985, pp. 462-71.) I disagree with McGee's conclusion about modus ponens, but think that his examples should be used to make some important points about assumptions, beliefs, ordinary conditionals, and valid inference. These points will show that no real counterexamples to modus ponens have been given.

McGee presents a number of very closely related examples, and he describes (on p. 462) the first of these, which is also the most attractive, in the following way. (The numbering (1)-(3) has been added by me.)

> Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised on the poll results believed, with good reason:
>
> (1) If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.
> (2) A Republican will win.
>
> Yet they did not have reason to believe
>
> (3) If it's not Reagan who wins, it will be Anderson.

McGee's conclusion in the light of examples of this type is that modus ponens is 'not strictly valid' (p. 462), and he expresses this point in the following way (p. 463):

> Sometimes the conclusion of an application of modus ponens is something we do not believe and should not believe, even though the premisses are propositions we believe very properly.

The first important point to notice is that McGee speaks of what we *believe* in the above quotation, and not of what we *assume*. In fact, he never speaks of modus ponens as the rule which allows us to infer a conclusion ψ from *assumptions* of the form ϕ and $\phi \Rightarrow \psi$ (or from assumptions these forms depend on), and yet this is how the rule is standardly stated in systems of natural deduction.[1] Actually (3) does validly follow given that we *assume* (1) and (2). Having a belief is not at all the same thing as making an assumption, and one respect in which they differ is that a belief *may* be suspended or set aside when an assumption by its very nature cannot be.

Suppose you believe that Reagan will win the election, and I ask you to consider what will happen if he does not win. You have no difficulty – you suspend or set to one side (in very informal terms) your belief about Reagan and certain other related beliefs, and then try to see what plausibly follows. But suppose now I ask you what will happen if Reagan does not win the election on the assumption that he will win it. This is puzzling because I am asking you, in effect, what follows from a contradiction. You do not immediately suspend or give up the assumption that Reagan will win the election, in order to decide what will happen if he does not win it, for what in that case would be the point of making the assumption? An assumption is a proposition we hang on to in circumstances like this, and to make the assumption is to agree for a time to hang on to the proposition.

Let us return to (3) above, and imagine that we are presented with this conditional on its own and asked whether we should believe it. To answer this question, we suppose that the antecedent of (3) is correct and, to avoid inconsistency, suspend some of our present beliefs, including, of course, our belief that Reagan will win the election. But by suspending that belief, we see that we no longer have any reason to believe (2) above, that a Republican will win (since we have presumably based this belief on our belief that Reagan will win). And as a result of these changes and what we know about the polls, we see that we should not believe (3), or more accurately the proposition which (3) expresses in this context.

---

[1] As McGee himself admits in a footnote (p. 462), modus ponens should really be seen as an inference involving *propositions* expressed by sentences of certain forms – a point relevant to what follows.

The matter is very different in a context in which we infer (3) from the assumptions (1) and (2) by modus ponens. Here we cannot 'suspend' (2) – in this context (2) is an assumption, and the rule modus ponens does not permit us to discharge its assumptions (modus ponens is not like conditional introduction in this respect). And so (3) does follow given the assumption (1) and (2), or more accurately the proposition *here* expressed by (3) does follow from the propositions expressed by the assumptions (1) and (2).

There are in addition serious arguments against any attempt like McGee's to define validity in terms of justified belief. Of course, there is the well known point that the premisses of a valid argument may *individually* have a higher probability than its conclusion, and some would argue for the analogous point about justification.[2] Although the probability of ϕ may be relatively high and also the probability of ψ, the probability of ϕ & ψ may be relatively low. Admittedly, the probability of the conjunction of the premisses cannot be higher than that of the conclusion in a valid inference, but some epistemologists have argued that degree of justification is not similarly closed under known valid inference. They claim that one may not be justified in believing a proposition ψ, even though one is justified in believing the single proposition ϕ and knows that ψ may be validly inferred from ϕ.[3] This claim of these epistemologists should be questioned, I think, but should not be ruled out as trivially false by the very definition of validity. Finally, how would validity be defined for inferences in which assumptions were discharged? We would not want to be faced with McGee-type 'counterexamples' to disjunction elimination and existential elimination because assumptions had been discharged, leaving conclusions apparently unsupported by justified beliefs alone.

All the points made so far here are relevant to a much wider range of examples than those discussed by McGee, as illustrated by this inference:

---

[2] This criticism of McGee is stressed by Walter Sinnott-Armstrong, James Moor, and Robert Fogelin, 'A Defence of Modus Ponens', *The Journal of Philosophy*, May 1986, pp. 296-300.

[3] The influential source of this view seems to be Fred Dretske, 'Epistemic Operators', *The Journal of Philosophy*, 1970, pp. 1007-23. It is closely related to the claim that knowledge is not closed under known logical implication – see also Robert Nozick, *Philosophical Explanations* (Harvard University Press, 1981), chap. 3.

(4) If the polls are right, then Reagan must win;
(5) The polls are right;
(6) Reagan must win.

Suppose we ask whether we should believe (6) above when we merely have good reason to believe, but do not assume, (4) and (5). Surely (6) may strike us as too strong a statement to be worthy of belief in this kind of case. If (5) expresses only a reasonable belief, and not an assumption, then we may not hold it constant in the set of possible situations we use to determine the acceptability of what is here expressed by (6). To suspend our belief in (5) is to be prepared to consider possibilities in which Reagan does not win, and so these possibilities determine what is expressed by (6) and the unacceptability of this proposition in this case. But if we go to a context in which we take (4) and (5) as assumptions, then we see that the relevant set of possible situations for judging the new use of (6) cannot have members in which either (4) or (5) is false. To assume (4) and (5) is to exclude such possibilities, and thus the 'narrower' proposition which (6) expresses in this context is true.

Some logicians would argue that there is a scope ambiguity in (4), and some of these would claim that (4) is only true (in the general circumstances described above) when it is given the logical form $\Box(\phi \Rightarrow \psi)$. With (4) in this form, of course, (4)-(6) would no longer appear to be an instance of modus ponens, and one would not need to ask whether (4) and (5) were assumptions. But this way of dealing with (4)-(6) as an apparent counterexample to modus ponens does not seem to be open to McGee, who appears to be committed to a hard line on giving logical forms to sentences in ordinary English. Towards the end of his paper he considers the claim that the logical form of (1) should be translated as $(\phi \,\&\, \psi) \Rightarrow \theta$ in a formal logic for the conditional. Clearly, if we only believed (1) in this form, then McGee could not argue that (1)-(3) is a counterexample to modus ponens. But he charges that it is unnatural to translate (1) in this way, and seems to take a hard line on translations in general in the following passage (p. 471):

> The selective use of unnatural translations is a powerful technique for improving the fit between the logic of natural language and the logic of a formal language. In fact, it is a little too powerful. One suspects that, if only one is sly enough in giving translations, one can enable almost any logic to survive almost any counterexample. What is needed is a systematic account of how to give translations. In the absence of such an account, the unnatural translations will seem likely merely an ad hoc device for evading counterexamples.

McGee does have a good point, to some extent, about the practice of some logicians in trying to give such translations, but he himself is in serious trouble. His account of validity in terms of justified belief and his strong comments on translation imply the very implausible result that there are many types of quite straightforward counterexamples to modus ponens. And we can bring out even more clearly what is wrong with these supposed counterexamples by considering the following modification of (1)-(3):

(7) If a Republican wins, then if he is not Reagan he will be Anderson;
(8) A Republican will win;
(9) If he is not Reagan, he will be Anderson.

The antecedent of (7) restricts the possibilities for the interpretation of the pronoun in its consequent. The second assumption (8) does the same job for the conclusion (9), and it would be a transparent mistake to try to interpret 'he' in some other way, in an attempt to show that (7)-(9) is invalid. McGee would make a mistake of this type if he thought of (8) as a relatively long-lasting mental state of justified belief outside of the context of this inference. He would not then see (8) as an assumption in an inference, determining in that context which proposition is expressed by (9).

An inference should be defined in terms of a relationship between assumptions and a conclusion, as is standard in logic. We should remember that the assumptions can restrict the relevant set of possibilities and so affect the propositions expressed under them, just as the antecedents can affect the propositions expressed by the consequents of conditionals. We must therefore be careful about the propositions expressed in inferences, particularly ones containing conditionals, if we wish to question their validity.[4] This is as true for (1)-(3) and (4)-(6) as it is for (7)-(9), and the point also applies to inferences of other forms, such as modus tollens, which can appear to have McGee-type 'counterexamples'. And when we use any of these valid inferences to extend our justified beliefs, we must be careful about which propositions we are to believe. Once these steps are taken, McGee-type 'counterexamples' to valid inference will disappear. Validity can be given a proper *semantic* definition in terms of propositions and truth preservation, and on that firm basis *epistemology* can legitimately ask the question of how justified belief is related to validity.[5]

Philosophy Division,
Sunderland Polytechnic,
Sunderland SR1 3SD

© D. E. OVER 1987

---

[4] This point can be used to support those who claim that (3) should not be interpreted in the same way as the consequent of (1). See, for example, E. J. Lowe, 'Not a Counterexample to Modus Ponens', *Analysis* 47.1, January 1987, pp. 44-7.

[5] For helpful discussion, I should like to thank Simon Blackburn, Dorothy Edgington, Hugh Mellor, Richard Spencer-Smith, and especially Wilfred Hodges and E. J. Lowe. Very special thanks are due to Kit Fine, who must be ultimately responsible for any good ideas in the paper.

FREE AND CLASSICAL COUNTERPARTS: RESPONSE TO LEWIS

By GRAEME FORBES

I

IN his recent book *On The Plurality of Worlds* (Blackwell, 1986; henceforth 'PW') David Lewis defends the view that there is no reasonable way of using counterpart theory to interpret the language of quantified modal logic so as to make it 'a well-behaved formal language, free of ambiguity ... free of devious semantic rules that work different ways in different cases ... [and also one which] affords a good regimentation of our ordinary modal thought' (PW p. 12). This conclusion is diametrically opposed to that of my 'Canonical counterpart theory' (ANALYSIS 42.1, January 1982, pp. 33–7, henceforth 'CCT', with a clarification in 'More on counterpart theory', ANALYSIS 43.3, June 1983, pp. 149–52); yet despite Lewis's criticisms, I am still convinced that the CCT system, which I called 'canonical' because of its mimicry of the canonical Kripke semantics for quantified S5, is the best implementation of Lewis's basic semantic ideas. In this paper, I will first state what is at dispute between Lewis and myself in as non-technical a way as possible (I assume the reader is familiar with the basics of and motivations for counterpart theory – see PW ch. 4.2 for Lewis's, *The Metaphysics of Modality*, Oxford 1985, ch. 7 for mine). Then in Part II, I will argue that the objections Lewis advances against CCT are without force. Finally, in Part III, I will try to show that Lewis's positive proposals are objectionable.

Suppose we agree that $\Box$ ($\Diamond$) can be understood as a universal (existential) quantifier over worlds. Then in useful terminology Lewis employs when he wants neutrality between counterpart-theoretic interpretations and the standard Kripkean one, we can say that for $\ulcorner \Box F\alpha \urcorner$ ($\ulcorner \Diamond F\alpha \urcorner$) to be true is for every (some) world to represent $\ulcorner F\alpha \urcorner$ as being the case. Characteristic of counterpart theory is its account of *what it is* for a world to represent $\ulcorner F\alpha \urcorner$ as being the case: a world represents *de re* by having *representatives* of the relevant objects with the appropriate properties, the representatives being called counterparts, each existing in exactly one world. So Lewis offers:

(1) w represents it as being the case that α is F iff α has a representative which exists at w and which is F at w.

More generally, if A is a complex sentence $\ulcorner \varphi(\alpha) \urcorner$ whose only name is 'α', and where, to avoid irrelevant complications, φ is non-modal, then we have

(2) w represents $\ulcorner \varphi(\alpha) \urcorner$ as being the case iff α has a representative which exists at w and which satisfies φ at w

in which (one would hope) the notion 'satisfies φ at w' can be given an uncontroversial recursive definition for non-modal φ. But in order to interpret claims of contingent existence properly, an exception to (2) must be made when φ is 'does not exist'. For by the standard recursive account, an object satisfies 'does not exist' at a world iff it does not satisfy 'exists' at that world. But from this and (2), we obtain

(3) w represents 'α does not exist' as being the case iff α has a representative which exists at w and which does not satisfy 'exists' at w

which is contradictory when 'exists' receives its intended interpretation. Thus no actual object is ever represented by any world as not existing; so no actual object exists contingently.

Lewis's solution is simply to make a special case (PW p. 11). When φ is 'does not exist', ignore (2) and say rather

(4) w represents $\ulcorner\varphi(\alpha)\urcorner$ as being the case iff α has no representative which exists at w

and this embracing of non-uniformity is the main step on his way to his negative conclusions about the language of quantified modal logic. But my CCT uses a strategy which *prima facie* preserves uniformity. The problem with Lewis's scheme is that an object's representative at a world perforce exists at that world. But can we not make sense of a relation 'ξ is a representative of ζ at w' which is free of existential commitment on the left? With such a 'free' counterpart relation, as opposed to Lewis's 'classical' one, we have

(5) w represents $\ulcorner\varphi(\alpha)\urcorner$ as being the case iff α has a representative at w which satisfies φ at w

and when the counterpart relation is free there is no problem with 'does not exist': a world represents α as not existing iff α has a representative at w which does not exist at w. For technical convenience, stipulate that α is its own representative at any world where it has no existing representative. Then $\ulcorner\Diamond\sim E(\alpha)\urcorner$ is easily satisfied, which seems to cut the ground from under Lewis's case for pessimism about the language of quantified modal logic.

II

Why does Lewis reject a system based on (5)? He has three objections to it, which I take in turn. First, he says, we cannot make sense of free counterparthood (PW p. 10): '... what can it mean to say that the counterpart is "at w" if not that, at w, the counterpart exists? ... We might just *say* it, and not mean anything by it. That is Forbes's solution ...' Most certainly it is not. If Lewis understands his own classical counterpart relation then he must understand my free one, since the latter is easily defined from the former with the aid of uncontroversial auxiliaries. Let $\mathcal{R}^*$ be the classical and $\mathcal{R}$ the free relation. Then

$$(6)\ \langle x, \alpha, w\rangle \in \mathcal{R} =_{\mathrm{df}} \text{either } E(x, w)\ \&\ \langle x, \alpha\rangle \in \mathcal{R}^*, \text{ or } \sim(\exists z)(E(z, w)\ \&\ \langle z, \alpha\rangle \in \mathcal{R}^*)\ \&\ x = \alpha.$$

I am not advancing any claim about which of $\mathcal{R}$ and $\mathcal{R}^*$ is 'genuinely' primitive relative to the other; the point is just that both mean something or neither do.

The condition for being a free counterpart is disjunctive. As I understand Lewis's second objection to free counterparthood, it is that a disjunctive condition for being a counterpart is no better than a disjunctive procedure for applying counterpart-theoretic interpretation, as embodied in (2) and (4). He writes (PW p.10 n.6): 'Forbes's solution ... amounts to a stipulation that there are two different ways that Humphrey ... can satisfy formulae *in absentia*. Where he has proper counterparts, he does it one way ... the ordinary vicarious way. Where he doesn't, he does it another way – just by not being there he satisfies "x does not exist".' What is satisfaction *in absentia*? Since the individuals of counterpart theory are worldbound, any satisfaction of a formula by an object at a world other than its own must be *in absentia*; α satisifes φ *in absentia* at w iff w represents α as satisfying φ ('Satisfaction *in absentia* is vicarious satisfaction' – PW p. 10); that is, iff a representative of α satisfies φ at w. But counterpart theory based on (5) gives an account of 'α satisfies φ at w' which works for *every* φ, and so, by these equivalences, gives an account of satisfaction *in absentia* which works for every φ; counterpart theory based on (2) and (4) evidently does not. So I reject the objection that it is *free* counterpart theory that gives different accounts of satisfaction *in absentia*. Lewis is conflating the question of how an object comes to be α's representative at any given world with the question of what role is played by representatives *vis-à-vis* representing that α is thus-and-so. I say there are two routes by which x can come to be α's representative, but a single mechanism by which interpretation works. Lewis discerns a single way to be a representative (though couldn't the similarity criteria defining $\mathcal{R}^*$ be disjunctive?) but different interpretative procedures. The advantage of my procedure is that it facilitates uniform interpretation of modal language.

Lewis's third objection is that the uniform interpretation CCT provides is not the one 'we would offhand expect' (PW p. 10). Assume for this discussion that α has at most one representative at any world (this is to finesse irrelevant disagreement about whether $\Box$ involves existential or universal quantification over counterparts). Then the 'problem' is that by (5), the truth of $\ulcorner\Box\varphi(\alpha)\urcorner$ requires that at every world α's representative satisfies φ, without requiring that the representative exist at the world. So suppose that we want to say that α has the property of being human as an essential property, which means, as Lewis agrees, that

(7) no possible world represents α as existing and also as not being human.

Then the difficulty with free counterparthood, allegedly, is that we cannot express 'α is essentially human' as $\ulcorner\Box\text{Human}(\alpha)\urcorner$, since this would require that at every world a representative of α satisfies 'is human', which asks more than (7) does and is false if (i) α is a contingent existent, and (ii) existence is necessary for being human. The difficulty for a proponent of (5) is then that he must 'discipline himself ... never to say "necessarily human" when he means "essentially human"; but instead always to say "necessarily such that it is human if it exists" ' (PW p. 12).

Why is this a *difficulty* for the proponent of (5)? Why would we 'offhand expect' to express 'α is essentially human' in Lewis's preferred way as $\ulcorner\Box\text{Human}(\alpha)\urcorner$ rather than $\ulcorner\Box(\text{Exists}(\alpha) \supset \text{Human}(\alpha))\urcorner$? I would have thought that offhand we would have no view about the matter, or even that we might offhand be mildly surprised that mention of existence disappears in Lewis's regimentation of 'α is essentially human' though it is explicit in the possible worlds explanation (7) of 'essentially'. Moreover, in using $\Box$ to express 'essentially', Lewis is giving it its so-called 'weak' reading, as made familiar by Kripke (see 'Identity and Necessity', in Munitz (ed.), *Identity and Individuation*, NYU Press 1971, p. 137). But it is a well known problem with weak necessity that it is expressively inadequate: we cannot use $\ulcorner\Box(\text{God exists})\urcorner$ to express what an orthodox theologian would mean by 'God exists of necessity', since the formula requires merely that God exist in all the worlds where He exists. Perhaps we should introduce an unstructured predicate 'exists of necessity' to express the theological claim, but on this approach, the pair of sentences 'God exists of necessity' and 'α is human of necessity' illustrate an ambiguity in 'of necessity', instead of, as the theologian would say, a clash of truth-values. Is this what one would offhand expect? Surely there is little to be said for these *ad hoc* manoeuvres when free counterpart theory affords a uniform treatment merely at the 'cost' of not treating 'necessarily' and 'essentially' as cognates (see further Martin Davies, 'Review of *Philosophical Papers Volume I* by David Lewis', *British Journal for Philosophy of Science* 37 (1980), pp. 130–4).

Two final points. First, the difficulties over expressive inadequacy which Lewis's proposals encounter demonstrate that the decision where to locate the 'disjunctive element' of the theory is far from the even trade-off Lewis's second objection implies it to be. For on the view that there are two ways to be a representative but just a single interpretative mechanism, no problem of expressive power arises. Secondly, although there is no way for Lewis to capture the meaning ascribed to $\Box$ by the standard clause in free counterpart theory, it is easy on the free semantics to express what Lewis means by $\Box$. The operator $\boxdot$ defined by

(8) $\ulcorner \boxdot\varphi(\alpha) \urcorner$ is true iff for every w, if x is α's representative at w and x exists at w, then x satisfies φ at w

expresses 'essentially'. Therefore my difference with Lewis is not just verbal, for I enjoy a generality he cannot match.

III

Define *property actualism* to be the view that (necessarily) no object x can possess an unstructured property or stand in an unstructured relation if x does not exist. Perhaps property actualism seems plausible, but it is not trivial to explain why or to deal with apparent counterexamples. Why should we be obliged to deny that Socrates belongs to {Socrates} at worlds where Socrates does not exist? And supposing that the existence of Socrates is necessary for the existence of propositions which are *de re* with respect to Socrates, can we not say that the proposition that Socrates does not exist is true (has the property of truth) at worlds where he does not exist, and where, consequently, that proposition does not exist?

Certainly, the property actualist has ways of dealing with these examples. For instance, he can admit an existentially unrestricted membership relation and an existentially unrestricted truth property on the grounds that they are not unstructured, but are rather definable from properties and relations in conformity with property actualism. But then he has to advance some definitions, and their adequacy is likely to be controversial. So there are substantial philosophical matters in dispute here.

Unfortunately, Lewis's account of counterpart theory legislates in advance in favour of property actualism, for as inspection of (1) shows, where R is atomic, every instance of

(9) $\Box(\forall x_1)\Box\ldots\Box(\forall x_n)\Box((Rx_1 \ldots x_n) \supset E(x_1) \,\&\, \ldots \,\&\, E(x_n))$

is translated into a theorem of his counterpart theory, and (9) is the linguistic version of property actualism. This is a drawback of Lewis's scheme which he himself must recognize, since he agrees with Kaplan that logic 'should *serve* philosophical ideologies, not constrain them' (PW p.105).

If there is this problem with his own approach, and if his case against CCT is as weak as I have made out, what motivates Lewis to take the line he does? It is because he wants to take advantage of an extremely simple account of what it is for a world to represent that such-and-such which is afforded him by his realism about worlds. That is, a world represents that certain objects have certain properties by having as (literal) *parts* representatives of those objects which have those properties. By contrast, he argues, views according to which worlds are maximal actually existing abstract objects have great difficulty in explaining representation (PW pp. 174–91). But while it would certainly be ironic if the attempt to combine realism about worlds with counterpart theory ran into difficulty over representation, I do not see that Lewis's realism would be compromised by complicating the account of representation in the minimal way of (5).

104, Social Science Building,
Tulane University, New Orleans,
Louisiana 70118, U.S.A.

© Graeme Forbes 1987

LEWIS ON PERDURANCE VERSUS ENDURANCE

By E. J. Lowe

David Lewis has recently presented a new argument (or perhaps only a new version of an old argument) to the conclusion that things which persist through change must do so by virtue of possessing *temporal parts* (see his *On the Plurality of Worlds*, Oxford: Blackwell, 1986, pp. 202 ff.). In his terminology, 'something *perdures* iff it persists by having different temporal parts, or stages, at different times, though no one part of it is wholly present at more than one time; whereas it *endures* iff it persists by being wholly present at more than one time' (p. 202). In these terms, his conclusion is then that ordinary changeable things like people and puddles (his examples) *perdure* rather than *endure*. This is a conclusion I very much want to resist, chiefly because I find the notion that such things *have* 'temporal parts' scarcely intelligible. Lewis suggests that this disqualifies people like me from entering the debate over perdurance versus endurance, because one cannot debate something one doesn't (or purports not to) understand. However, it isn't that I don't understand the notion of a temporal part as such – I understand it perfectly well as applied to *events* or *processes* – it's just that I dispute its applicability to entities in the category of *continuants*. Anyway, I *shall* enter the debate. I should just say, though, that in seeking to resist Lewis's conclusion, I don't mean necessarily to endorse the suggestion that things like people or puddles do in fact *endure*, as *Lewis* defines this term. This is because I can find no useful application for the notion of such a thing being 'wholly present' at a time, any more than I can for that of its being 'partially non-present' (in the sense, of course, of having earlier or later parts). (Certainly, I don't find talk of a thing's being 'wholly present' at any time at which it exists a perspicuous way of denying that the notion of temporal parts is applicable to it.) All I am concerned to do, then, is to resist the contention that such things 'perdure'.

Lewis's argument turns on what he calls the 'problem of temporary intrinsics':

> Persisting things change their intrinsic properties. For instance shape: when I sit, I have a bent shape; when I stand, I have a straightened shape. Both shapes are temporary intrinsic properties; I have them only some of the time. How is such change possible? (pp. 203–4)

Lewis says that he knows of only three possible solutions. The first is to deny that shapes are genuine intrinsic properties, saying instead that they are 'disguised relations, which an enduring thing may bear to times' (p. 204). I can agree with Lewis that this solution (which would seem to represent a thing's *shape* at a time as being somewhat akin to its *age* at a time) is untenable. The second is to say that 'the only intrinsic properties of a thing are those it has at the present moment' (ibid.), the implication being a denial of the reality of both the past and the future. Again I can agree that this solution is untenable. Lewis's third solution is the one he favours: 'the different shapes ... belong to different things' (ibid.), these different things being precisely different *temporal parts* of the changeable, persisting thing, which therefore *perdures* rather than *endures*. Thus one temporal part of me has a bent shape, another a straightened shape.

However, these three candidate solutions do not exhaust the field. There is another candidate which, in my view, squares far better with common sense than any so far mentioned. To simplify its presentation, I shall consider first not the changing shape of a *person*, but rather the changing shape of a rather simpler object, O, consisting of two flat, rigid wooden planks P_1 and P_2 joined endwise by a hinge. At one time O may have a bent shape and at another it may have a straightened shape, just as Lewis says of himself. How is this possible? The answer is quite obvious: at one time P_1 and P_2 are at an angle to one another and at another they are in alignment. O's changing *shape* is to be explained in terms of the changing spatial *relations* between its constituent parts, P_1 and P_2, whose own shapes (we have been assuming) are unchanging. O's shape at any given time is *not*, then, a relation (not, in particular, a 'disguised' temporal relation), but it does 'supervene' upon the shapes and spatial relations of its constituents at that time. And as it is with O, so it is more complicatedly with more complex objects like people.

This explanation of O's changing shape presupposed, it is true, that P_1 and P_2 were rigid, so that *their* shapes were unchanging. But there is no reason why we shouldn't allow them too to change in shape, explaining this in turn in terms of the shapes and changing spatial relations of *their* constituent parts. (Notice that when I speak here of 'parts' of changeable objects, I mean this in the ordinary, common-sense way in which P_1 and P_2 are *parts* of *O*; I am not speaking of 'temporal' parts.) Now, obviously, we have set out here upon a *regress* – but it isn't at all obvious that it must be a *vicious infinite* regress. The regress can perfectly well be terminated at a level of *fundamental particles*, which have all their intrinsic properties *unchangeably*. Classical atoms had their *shape* unchangeably; the fundamental particles of modern physics – quarks and so on – don't of course have shapes at all, in any literal sense, though they do have other intrinsic properties (like spin, charge and 'colour') unchangeably. And ultimately, or so we are led to believe, it is in terms of such properties of such particles, together with the relations of such particles to one another, that the whole macroscopic physical realm is to be explained. (Lewis himself concedes, without seeing the significance of so doing, that fundamental particles may well 'have no accidental intrinsic properties' (p. 205, note 6).) In short: modern physics offers us a solution to the problem of change which renders superfluous Lewis's solution in terms of temporal parts. Indeed, in comparison with the scientific solution, Lewis's 'solution' has all the hallmarks of the sort of pseudo-explanation that used to give philosophy a bad name.

University of Durham,
50 Old Elvet,
Durham DH1 3HN.

© E. J. LOWE 1987

ON KNOWLEDGE OF THE UNKNOWABLE

By TIMOTHY WILLIAMSON

IF it is an unknown truth that *p*, it is an unknowable truth that it is an unknown truth that *p* (cf F. B. Fitch, 'A logical analysis of some value concepts', *The Journal of Symbolic Logic* 28, 1963, pp. 135–42, at p. 138). It follows, by classical logic, that if all truths are knowable then all truths are known. This hardish fact makes life difficult for the verificationist who wishes to assert that all truths are knowable, but to deny that all truths are known. He might try rejecting classical logic (cf. my 'Intuitionism disproved?', ANALYSIS 42.4, October, 1982, pp. 203–7). Dorothy Edgington has recently suggested a different way out (cf. 'The paradox of knowability', *Mind* 94, 1985, pp. 557–68). She admits that there are unknowable truths in the sense of Fitch's argument, but interestingly revises the verificationist principle, to give it this form: if, in a situation *s*, it is a truth that *p*, then there is a possible situation *s′* in which it is known that, in *s*, *p*. Although the truth (for some values of '*p*') that, in *s*, it is an unknown truth that *p* cannot be known in *s*, it might be known in some possible situation other than *s*. I have argued elsewhere that this account faces difficulties of principle in explaining how the knower in *s′* can pick out the situation *s* in thought with adequate specificity (cf. 'On the paradox of knowability', *Mind* forthcoming). I waive those objections here, to examine Edgington's application of her principle to a particular case of knowledge in one situation that something is an unknown truth in another. I shall argue that her account does not generalize to the full range of cases which it needs to cover.

Suppose that, in the actual situation *s*, it is an unknown truth that it rained last night. Nevertheless, someone *could have* known that it rained last night. How would this provide knowledge that, in *s*, it is an unknown truth that it rained last night? According to Edgington, 'the possible knower that it rained last night has to know that it still would have rained, had neither he nor anyone else noticed. And any possible knower that it rained does, or at least can, know this' (p. 567). Such an account obviously does not apply to cases in which the corresponding counterfactual is false, but, Edgington says,

> If there are truths which fail to satisfy the principle
>
> (1) *p* would still have been true, had no one known that *p*,
>
> that I am in pain, for example, then they satisfy the principle,
>
> (2) If *p*, then someone knows that *p*
>
> and, *a fortiori*, they satisfy [the verificationist principle] (ibid., my numbering).

Evidently, the above account of knowledge that, in *s*, it is an unknown truth that *p* presupposes that it is knowable that *p* (in the ordinary sense, in which Edgington concedes that not all truths are knowable). It therefore does not cover a case of the following kind. Let '*p*' abbreviate a description of all the neurophysiological events which ever actually take place in any brain, conjoined with the proposition that no other neurophysiological events take place in any brain. It is an unknown truth that *p*. However, it does not even seem knowable by creatures such as us that *p*, for such knowledge would have to be recorded in a brain by different states from those which record actual knowledge, making it false that *p*. Thus Edgington's account does not show how there could be knowledge that, in *s*, *p*, for this value of '*p*'. This limitation noted, one can look at those cases where it is known in some situation *s′* that *p*.

Suppose that (1) is known in *s′* to be true by someone who also knows that *p*. Edgington's argument requires this to constitute knowledge that, in *s*, it is an unknown truth that *p*. Now clearly, in *s′*, (1) could not tell one that it would still have been true that *p* in *s*, unless *s* would or might have obtained, had no one known that *p*; otherwise, its being true that *p* in *s* would not be a necessary condition for the truth of (1). For the truth of a counterfactual does not require its consequent to be true in all possible situations in which its antecedent is, but only in those which are most similar to the situation at which the counterfactual is being evaluated, for some appropriate respect of similarity. Thus Edgington's argument requires the following counterfactual *not* to hold in *s'*:

(3) *s* would not have obtained, had no one known that *p*

For this reason, *s'* should be chosen so as to differ as little as possible from *s*, given that it is known that *p* in *s'* but not in *s* (and that (1) is known in *s'* but not in *s* – this may force further, unwanted differences between *s'* and *s*). Even when *s'* is chosen in this way, however, (3) often *will* hold in *s'*. For example, suppose that in *s* a flower bloomed on this spot at time *t* (=*p*), where *t* is a past time at which there were no relevant thinking beings. It is an unknown truth in *s* that *p*. Nevertheless, it could have been known that *p*; in some *s'*, human beings evolve earlier and see a flower here at *t*. Their counterfactual thought (1), that a flower would still have bloomed here at *t* had no one known that it did, is true. However, it is true simply because the flower would still have bloomed here, had they not gone for a walk in the forest that day; for situations of that kind, in which the general evolution of the human race up to *t* is more or less the same as in *s'*, are the most relevantly similar ones to *s'* in which the antecedent of (1) is true. In other words, (3) holds in *s'*, for, in *s*, the human race has not even begun to evolve at *t*. Hence, as argued above, knowledge in *s'* of (1) does not give knowledge of actuality. Note that it is not even clearly true in *s'* that a flower would have bloomed here at *t* had the human race not evolved until it did in *s*; the most relevantly similar situations to *s'* in which it does evolve later might be ones in which the absence of human beings altered the ecology of the area in such a way that a flower did *not* bloom here at *t*.

Thus, even if someone can know both (1) and that *p*, Edgington's account of knowledge that, in *s*, it is an unknown truth that *p* breaks down for some cases. Her other line of argument in the quoted passage is intended to apply when (1) is false. There is thus another lacuna: what if (1) were true but unknowable, or at least not knowable by whoever knew that *p* (in the sense in which Edgington concedes that not all truths are knowable)? However, instead of pressing this point, I shall examine Edgington's claim that if (1) is false then (2) is true, in which case the verificationist has no problem.

Note first that the difficulty arises when (1) is false in *s'*, the situation which would have obtained had it been known that *p*. If the falsehood of (1) entails the truth of (2), then presumably what follows is that (2) is true in *s'*, which is no surprise since the consequent of (2) is true by hypothesis in s'; but since the truth of (2) in s' does not seem to rule out the possibility that it is an unknown truth in s that p, no way seems to have been provided of knowing either in s' or in s that, in s, it is an unknown truth that p. Thus the revised verificationist principle is still under threat. Note that it would be irrelevant to try to argue to the truth of (2) in s from the falsity of (1) in s, for, by hypothesis, both the antecedent and the consequent of (1) are true in s. Thus it would have to be claimed that the falsity of (1) in s' entails the truth of (2) in s. But there is no reason why such an entailment should hold. For example, suppose that it is an unknown truth in s that exactly n lists of a hundred or more items were completed in Carthage from its foundation to its destruction ($=p$; I assume that $100<n$). The corresponding s' is presumably a situation in which a list of all the n relevant lists, including itself, is completed just as Carthage is about to be destroyed. Let s'' be a situation as relevantly similar to s' as is compatible with its not being known in s'' that p; s'' is presumably a situation in which the list of lists is not completed, and no lists are completed which were not completed in s'. Thus at most $n-1$ relevant lists are completed in s'', so in s'' it is not the case that p. Hence, by the choice of s'', (1) is false in s'. Nevertheless, (2) is not true in s – the Carthaginians did not actually make such a list of lists. In this example, it might be claimed that a modified version of Edgington's argument would go through, in which s' would be chosen as a situation in which the list of lists is completed and contains $n+1$ items including itself; in s', it would not be the case that p, but it would be known that it would have been the case that p had it not been known that exactly $n+1$ relevant lists were completed; this might be claimed to be knowledge of s. However, such a move is *ad hoc*: there is no longer any general way of finding a situation s' of the right sort, nor any reason to think that there is any such situation. Moreover, there are more recalcitrant examples. Suppose that our best physical theory tells us that one can know what state a ?-particle is in only by interacting with it in ways which unpredictably change its state. Suppose also that it is an unknown truth in s that ?-particle z is in state k ($=p$). Then (1) is not true in the corresponding s', but (2) is not true in s. Of course, a verificationist might deny that there can be *unknown* truths in cases of this kind: such a claim represents, however, a sharp increase in the price to be paid for the dubious advantages of verificationism. Certainly, there is no reason to regard the falsity of (1) in the relevant situation as any general guarantee of the actual truth of (2).

Edgington has offered a model of non-actual knowledge that, in the actual situation, it is an unknown truth at p. I have argued that, even if one accepts the model for the simple examples which she gives, it seems to break down for more complicated ones. Nor is there any obvious prospect of adapting the model to these cases. The model therefore provides no reason to think that the revised verificationist principle, which requires such knowledge, is universally true. Edgington also discusses weaker forms of verificationism, in which the concept of knowledge is replaced by that of reasonable belief, or of evidence; the arguments of this paper apply just as much to them. If revised verificationism lacks independent plausibility, Edgington would probably wish to rest her defence of it on an appeal to considerations about our grasp of concepts (p. 557). It would be interesting to have a detailed account of what it is about such considerations in virtue of which they lead to revised verificationism rather than to the fundamentalist principle, in the sense in which Edgington rejects it, that all truths are knowable.

Trinity College,
Dublin 2, Ireland

© TIMOTHY WILLIAMSON 1987

SALMON VERSUS KRIPKE ON THE *A PRIORI*

By RALPH KENNEDY

IN *Naming and Necessity* (Harvard, 1980) Saul Kripke wrote: 'The case of fixing the reference of "one meter" is a very clear example in which someone, just because he fixed the reference in this way, can in some sense know *a priori* that the length of this stick is a meter without regarding it as a necessary truth' (p. 63). I.e., if someone fixed the reference of 'one metre' by stipulating that it was to be the length at a time t of a certain stick S, then that person could know, *a priori*, that stick S was exactly one metre long at t, even though the proposition that stick S was a metre long at t is not a necessary truth. I will refer to this claim as 'Kripke's Claim'.

Nathan Salmon, in his recent book *Frege's Puzzle* (Bradford, 1986), rejects Kripke's Claim. At the same time he offers an explanation for why such a claim might have been made. The information that the sentence 'Stick S, if it exists, is exactly one metre long at t' is true 'may indeed be a priori for the agent in Kripke's example, and this apparent fact seems to be the source of Kripke's contention that the (content of) the sentence is itself a priori' (p. 141). The former information is pragmatically imparted by typical utterances of the sentence, but it is not what the sentence says: it is not the semantic content of the sentence. If one is not careful to bear this distinction between semantic contents and pragmatic impartations in mind, one may be led to conclude, with Kripke, that the introducer of 'metre' knows *a priori* that the stick is one metre long at t simply because it is so plausible that the introducer knows *a priori* that a certain sentence with this content is true (p. 141). But if one does bear the distinction in mind it will presumably be evident that the actual information *content* of the sentence is a very 'paradigm of a posteriori knowledge', of knowledge 'gained ultimately by measurement' (p. 142).

My aim in this paper is to call attention to the fact that Salmon could have given a stronger argument than he did against Kripke's Claim. Specifically, the theory Salmon develops in his book appears to contain the resources for a rather striking *reductio* of Kripke's Claim, leaving one to wonder why Salmon settles for doing little more than asserting that Kripke's Claim is implausible.

Salmon's theory is far too complex to be summarized here. For us, its salient feature is that it treats any two sentences (of a given language) that differ only in that where one has a simple singular term T the other has a co-referential simple singular term T' as semantically identical: they have the same 'program', which means that with respect to any context, time and possible world, they 'encode' the same proposition (see chapter 2, *passim*). Since Salmon's theory takes propositions to be the objects of knowledge and belief (p. 133), it follows that anyone who believes that Venus is Venus thereby believes that Hesperus is Phosphorus: the proposition that Venus is Venus and the proposition that Hesperus is Phosphorus are one and the same. (In fact, the proposition that if Hesperus exists then it is Phosphorus is 'fully knowable, with complete certainty, by reason alone', p. 137.)

Consider, then, with reference to this (part of the) theory the following story. Someone we'll call 'Claudia' has just found a stick that looks in every way like a measuring stick except for lacking any numerals or other markings. The stick is in fact exactly one metre long. Claudia says to herself: 'This is certainly not a yard stick: it is too long. Perhaps it is a metre stick. No, I'm sure it's not long enough for that.' It would seem safe to say that Claudia – assuming that the language she is speaking is, as it seems to be, English – does not know that the length of the stick is exactly a metre.

What I wish to demonstrate is that, on Salmon's theory, if we were to accept Kripke's Claim about the possibilities of *a priori* knowledge for introducers of terms, we would be forced to accept the obviously absurd conclusion that Claudia could change her epistemic situation, merely by engaging in a sort of baptism ritual, to one of knowing (and *a priori* at that) that the length of the stick is exactly one metre. Here is what she could do. She could say: 'I think I'll call the length of this stick, which is certainly more than a yard and less than a metre, a "schmoo".' If Kripke's Claim were true, Claudia would now know that the length of the stick, if it exists, is exactly one schmoo, and she would know this *a priori*. But the proposition that the length of the stick is a schmoo is, on Salmon's theory, the *same* as the proposition that the length of the stick is a metre, since 'schmoo' and 'metre' are simple co-referential singular terms. Hence, since the objects of knowledge and belief are propositions, Claudia would know *a priori* that the length of the stick, if it exists, is exactly one metre, and she would have come to be in this epistemic state (from one in which she definitely did *not* know that that length was a metre) merely by engaging in the baptismal act. Surely few claims vie with this one for absurdity, or so it would seem.

(Doubts may well arise about the claim that 'schmoo' and 'metre' are simple singular terms. But Salmon appears to count as 'simple' just those terms which consist of one word and are not 'generated by semantic stipulation, for example, the stipulation that a certain apparently simple expression is to have precisely the same information value as a certain compound one' (p.138). 'Metre' and 'schmoo' would seem thus to qualify as 'simple'. Are they *singular*? Salmon notes that 'metre' is a common noun and hence 'apparently a general term rather than a singular term' (p. 178, n. 11); nevertheless, he proposes for present purposes to treat 'unit measurement terms such as "meter" as singular terms that combine with number terms in combination with pluralization ("three meters") to form new singular terms' (ibid.). Thus, at least within the terms of Salmon's theory, we are on firm ground in treating both 'metre' and 'schmoo' as both simple and singular.)

I turn now to an objection to my argument. Could Salmon really say that Claudia *doesn't* know that the length of the stick is a metre? The proposition that Claudia knows that the length of the stick is a metre seemed absurd to us because Claudia acted, talked and reasoned as if she did *not* know or even believe this. But it might be urged that such facts about the way Claudia acted, talked or reasoned cannot be appealed to, on Salmon's theory, to justify the claim that Claudia does not know the length to be one metre. After all, she may also have acted, talked and reasoned as if she did not know or believe that Hesperus was Phosphorus; but this would not have counted in the slightest against the claim that she knows that Hesperus is Phosphorus, on Salmon's account.

It is clear that Salmon needs to be able to say the following two things if our proposed *reductio* is to work: (1) before the dubbing, Claudia doesn't know that the length is a metre; (2) after the dubbing, she does – if Kripke's Claim is assumed. These two claims together generate what I take to be the obvious absurdity that the act of dubbing suffices for Claudia to pass from not knowing that the length is a metre to knowing (*a priori*) that it is a metre. But if the objection is correct, Salmon would have no right to assert (1).

I think, however, that the objection involves an incorrect interpretation of Salmon's refusal to treat anything Claudia might say or do as evidence that she does not know that Hesperus, if it exists, is Phosphorus. This refusal is motivated by specific theoretical reasons that apply only to specific sorts of cases. Salmon's theory implies that Claudia, whatever she may say or do, does know the proposition in question. Moreover his theory provides, as we have seen, an explanation of the apparent evidence to the contrary. The case of Claudia's knowing or not knowing that the length of the stick is a metre is quite different. There is no Salmon-theoretical reason for supposing that Claudia *must* know (or even believe) that the length is a metre. (Since 'the length of stick S' is not a simple term, Salmonian theory does not deem the proposition that the length of stick S is a metre to be the same as the proposition that a metre is a metre, which might seem plausibly attributed to Claudia.) Hence, there is no question of Salmon being forced to find a reinterpretation of any apparent evidence to the contrary. And it is unclear that Salmonian theory could support such a reinterpretation. It's hard to imagine better evidence than that built into our story against the claim that Claudia knows that the length of the stick is a metre.

I conclude then that Salmon would be perfectly within his rights to regard Claudia's behaviour as extremely strong, if not conclusive, evidence that she does not know, prior to the dubbing, that the length of the stick is a metre. Salmon is thus in a position to assert that Kripke's Claim implies, against the background of Salmonian theory, the following absurdity *not* implied by Salmonian theory alone: (1) 'by virtue of performing an act of dubbing Claudia goes from a state of not knowing at all that stick S, if it exists, is one metre long at *t* to a state of knowing this *a priori.*'

This leaves us with the question: have we really added anything to Salmon's reasons for rejecting Kripke's Claim? Our absurdity really doesn't *look* all that different from Kripke's Claim itself, i.e.: (2) 'by virtue of performing an act of dubbing Kripke's dubber goes from a state of not knowing at all that stick *S*, if it exists, is one metre long at *t* to a state of knowing this *a priori*'. Yet despite the virtually identical wordings, I would argue that (1) is much more *vividly* absurd than is (2), mainly for the following two reasons: first, 'metre' is imagined in Kripke's example as being a *new term*, and it may well be felt that an introducer of a new term has certain freedoms and opportunities with respect to that term that Claudia cannot have with respect to 'metre', which is imagined in my example as being already established; second, we have what almost everyone would admit to be virtually conclusive evidence, even when judged from the standpoint of Salmon's theory, that Claudia does not know, either before or after the dubbing, that the length of the stick is a metre. Nothing similar is available in the case of Kripke's dubber.[1]

Wake Forest University,
Box 7332, Winston-Salem,
North Carolina 27109, U.S.A.

© RALPH KENNEDY 1987

---

[1] I am grateful to an *Analysis* referee for comments on an earlier version.

ON LEMKE'S DEFENCE OF A CAUSAL BASING REQUIREMENT

By JONATHAN L. KVANVIG

LEMKE has recently taken issue (see ANALYSIS 46.3, June 1986, pp. 138–44) with my claim that no counterfactual causal account of the basing relation is plausible (see ANALYSIS 45.3, June 1985, pp. 153–8). Intuitively, a counterfactual causal account claims that belief is based on evidence if and only if the evidence either causes the belief or would have caused it had the actual cause been absent. This intuitive formulation accounts only for counterfactual causes of level one: events which would have been a cause had only the actual cause been absent. As I argued, there is as much support for allowing counterfactual causes having a higher cardinality: events which would have been a cause had the actual cause and some other counterfactual causes been absent.

This latter fact raises special problems of formulation for the counterfactual causal theory of the basing relation. One is tempted to claim the following: were the belief (*Sbi*) and the appreciation of the evidence (*Sbe*) to occur and were nothing else to have caused the belief (*Sbi*) then the appreciation of the evidence would have caused the belief. My objection in the earlier cited article was that this claim is a necessary truth unless it is possible that beliefs be uncaused. I concluded that the counterfactual causal theory cannot succeed because it must countenance counterfactual causes of a higher cardinality than level one, but it cannot do this without trivializing the theory.

Lemke claims that my objection to the generalized form of the counterfactual theory, that form which allows for counterfactual causes of a higher cardinality than level one, is misguided. First, he believes that the generalized form does not contain a counterfactual which is close to being a necessary truth; and second, he believes that there is a more plausible formulation of the generalized form of the counterfactual theory than the one I attacked. I believe he is wrong on both counts, and shall explain why.

I

The first of Lemke's two major claims is that the objection I gave to the generalized form of the counterfactual causal theory is inadequate. The objection I gave attempted to show that the generalized counterfactual of the counterfactual causal theory is too close to being a necessary truth, and hence that any counterfactual theory which employs that generalized counterfactual fails to impose any substantive requirements regarding basing beliefs on evidence. The proposed generalized counterfactual is:

(G1) If *Sbc* were not to cause *Sbi* and both *Sbe* and *Sbi* were still to occur and nothing else were to cause *Sbi*, then *Sbe* would cause *Sbi*.

Lemke claims that this is not as close to being a necessary truth as I claimed.

> Initially, it may be suspected that Kvanvig has simply confused the notions of counterfactual cause and logically possible cause. For the former implies the latter but not vice versa. The above condition (G1) seems to be a necessary truth only if we take 'nothing else' as 'no other logically possible cause'. Instead, if we take it as a more restrictive 'no other counterfactual cause', then its truth seems to be a contingent matter (p. 142).

Lemke then attempts to distinguish counterfactual causes from logically possible causes. He begins by partially defining set membership for counterfactual causes as follows:

(CC) where *Sbc* causes *Sbi*: (p) (if *Sbp* is a member of set C of counterfactual causes of *Sbi*, then *Sbp* does not cause *Sbi* and if *Sbc* and all members prior to *Sbp* (if any) were not to cause *Sbi* but *Sbp* and *Sbi* were still to occur, then *Sbp* would cause *Sbi*) (p. 142).

Lemke believes that, given (CC) and the notion of joint causal overdetermination, he can show that not every logically possible cause is a counterfactual cause. Two events are joint causal overdeterminants of a third event just in case both of the first two events cause the third event (p. 142). (I suspect Lemke means here to require the stronger claim that both the first two events are individually causally sufficient for the third event.) Lemke's argument proceeds by citing two features concerning joint overdetermination which he believes distinguish logically possible causes from counterfactual causes. Further, he holds that these features distinguish logically possible causes from counterfactual causes in such a way that not every logically possible cause is a counterfactual cause. I shall argue that Lemke has not shown that either of these claims is true.

The first feature Lemke explains as follows. '(A)ny merely logically possible cause, e.g. *Sbl*, can be a joint causal overdeterminant with the actual cause, *Sbc*, such that both cause *Sbi*. However, counterfactual causes cannot be a joint causal overdeterminant with the actual cause or any prior member(s)' (p. 142). The first claim in this argument is ambiguous. Lemke might mean that some possible causes are also actual causes, whereas no counterfactual cause can be an actual cause. This claim is true but irrelevant to the present discussion. All my original argument needs to show is that every non-actual logically possible cause is a counterfactual cause.

Alternatively, Lemke might mean this to be a modal claim, in which case it is not clearly true. As a modal claim, it asserts that there is some possible world in which both *Sbl* and *Sbc* cause *Sbi*, but that there is no possible world in which *Sbq* is a counterfactual cause relative to the actual world and yet in which both *Sbq* and *Sbc* cause *Sbi*. The first conjunct is true, but the second is true only if the actual cause is always necessarily 'prior to' every counterfactual cause. Lemke does not clarify the notion of priority appealed to in (CC), but on one plausible understanding of priority, this claim is false. Suppose we say that one cause (counterfactual or otherwise) is prior to another just in case the causal efficacy of the first must be absent in order for the second to be a cause. Given this understanding, nothing in (CC) rules out the possibility of there being worlds in which a counterfactual cause can be a joint overdeterminant with the actual cause.

So Lemke's first attempt fails to find a characteristic which distinguishes logically possible from counterfactual causes. Lemke's second attempt is cryptically stated as follows: 'if *Sbp* has a prior member, *Sbq*, then if *Sbq* were to cause *Sbi*, then *Sbp* would remain a counterfactual cause and as such could not be a joint causal overdeterminant (with *Sbq*) of *Sbi*' (p. 143). The argument here seems to be that it is possible that *Sbl* (the merely logically possible cause) and *Sbq* be joint causal overdeterminants even if *Sbq* is prior to *Sbp*, but that if *Sbq* is prior to *Sbp*, then it is not possible that *Sbp* and *Sbq* be joint causal overdeterminants. Hence, *Sbp* and *Sbl*, as arbitrary examples of counterfactual and logically possible causes, have different properties.

Nothing in (CC) entails that *Sbp* and *Sbq* cannot be joint causal overdeterminants even if we assume that *Sbq* is prior to *Sbp*. (CC) requires that a certain counterfactual be true. On the standard semantics, we evaluate that counterfactual by finding close worlds in which the antecedent is satisfied. For simplicity, let us assume that the order of priority in question is as follows: *Sbc* is prior to *Sbq*, which is prior to *Sbp*. Then, for (CC) to be true, the closest world to the actual world (if there is one) in which neither *Sbc* or *Sbq* causes *Sbi* must be such that, in it, *Sbp* causes *Sbi*. Nothing in this claim implies that there is not some other world, quite remote from the actual world and this closest world, in which *Sbp* and *Sbq* are joint causal overdeterminants of *Sbi*.

So neither of Lemke's attempts to distinguish logically possible from counterfactual causes are successful. On the other hand, suppose that Lemke is right regarding the first claim. I.e., suppose that logically possible causes can be overdeterminants but that counterfactual causes cannot. Does it follow that some logically possible causes are not counterfactual causes? No, it does not. At best, what Lemke shows is that there are some worlds in which some cause, a member of a pair of joint overdeterminants, is not a counterfactual cause relative to our world. It does not follow that that same cause, which is one of a pair of joint overdeterminants in one world, is not a cause in another world which counts as a counterfactual cause relative to our world. One and the same event can have the following properties: it can be one of a pair of overdeterminants in one world, and a counterfactual cause as well (in virtue of causing a particular event in another world). At best, all Lemke shows is that the *concepts* of a logically possible cause and a counterfactual cause differ; from that it does not follow that some logically possible causes are not counterfactual causes.

One final argument for thinking that all logically possible causes are counterfactual causes. On any plausible understanding of priority, every counterfactual cause will be prior to any logically possible cause which is not a counterfactual cause. This fact allows for a quick reductio argument against any supposed distinction between logically possible and counterfactual causes. For, if all counterfactual causes are prior to any merely logically possible cause (which is not a counterfactual cause), then the following counterfactual will always be true: if the actual cause of belief were absent and no counterfactual cause were to cause the belief, and some merely logically possible cause (which supposedly is not a counterfactual cause) and the belief were to obtain, then the logically possible cause would have caused the belief. The only way for this claim to be false would be for it to be possible for the belief in question to be uncaused, and that is a possibility we can ignore here for reasons given in my earlier article. But if that counterfactual is true, then the supposition that the merely logically possible cause in question is not a counterfactual cause is false.

II

Lemke's claims, however, are not limited to an attack on an argument I gave against the counterfactual causal theory, for Lemke's second major point takes us beyond the claims contained in my discussion of that theory. This point claims that there is a more adequate specification of the counterfactual in question than the one I attacked. He begins by noting that Lehrer's original gypsy lawyer case (see *Journal of Philosophy*, vol. 67, 1970, pp. 841-55) does not constitute a counterexample to

(1) if *Sbc* (the actual cause of *Sbi*) were not to cause *Sbi* and both *Sbe* and *Sbi* were still to occur, then *Sbe* would cause *Sbi*.

In Lehrer's original case, the lawyer's reading of the tarot cards is causally responsible both for the belief and for the lawyer's reexamination of the evidence (which he comes to see conclusively establishes his client's innocence). Further, in that case, if the cards were no longer to cause the gypsy's belief in his client's innocence, he would no longer have the belief that his client is innocent. Lemke believes this fact shows that (1) is not the best way to formulate the counterfactual theory. He then goes on to claim that a better formulation is:

(2) if *Sbc* were not to cause *Sbi* and *Sbe* were still to occur and nothing else were to cause *Sbi*, then *Sbe* would cause *Sbi*.

Lemke then explains why (2) is superior to (1).

> If this antecedent (the antecedent of (2)) is satisfied, then it remains an open question whether *Sbe* would cause *Sbi*. Kvanvig's claim that it *must* do so depends crucially upon the stipulation in the antecedent that *Sbi* still occur. Once this is removed, the difficulty vanishes (p. 143).

I originally claimed that (G1), the generalized form of (1), is a trivial truth. Lemke claims that generalizing the counterfactual theory trivializes it only if the generalization is (G1) rather than (2). In particular, the trivialization occurs because (1) contains the requirement that the belief in question occur in the counterfactual circumstances in question. If we drop this requirement, as (2) does, we will not be close to having a necessary truth, as I claimed we have if we generalize (1) to (G1).

This formulation of the counterfactual causal theory does avoid my trivialization objection. Nonetheless, it is not a more adequate formulation than the one I attacked, for (2) can be shown to be false by a very minor alteration of Lehrer's original counterexample. All that we need to do is to add to Lehrer's case the claim that if the tarot cards had not had their impact, something else would have occurred which would have resulted in the lawyer properly appreciating the force of the evidence and yet being unable to form the belief that his client is innocent. Suppose, for example, that the lawyer would have read the newspaper had he not consulted the tarot cards, would have noted certain inconsistencies, and would have re-examined the evidence and come to see its force because of his noting of these inconsistencies. Further, though, suppose that the article he read was partially editorial, and advised all parties concerned with the case to suspend belief because it would foster a more just decision. The lawyer is convinced by this argument, and so, in spite of seeing the force of the evidence, refuses to form the belief that his client is innocent.

The addition of these facts to the case in no way affects the intuition that the lawyer, at present, knows that his client is innocent. Note further that if a counterfactual causal theorist wishes to deny that knowledge is present in gypsy-lawyer-like cases, he removes the motivation for accepting a counterfactual causal theory. For, to be moved to accept such a theory, one must first hold that there are cases of knowledge in which the belief is not causally sustained by the evidence. So a counterfactual causal theorist will be in quite a predicament if he denies that gypsy-lawyer-like cases are cases of knowledge. Yet, if it is granted that knowledge obtains in this case, we then have a counterexample to (2), for in this case, the actual cause is absent, the appreciation of the evidence is present, but that appreciation does not cause the belief in question precisely because the belief is absent.

Note that (1) fares better regarding this case, for it requires that the belief be present in any relevant counterfactual circumstance. Further, this requirement is eminently plausible. If counterfactual circumstances are allowed in which the belief is absent, there will be a multitude of cases like the altered gypsy lawyer case above in which Lemke's counterfactual causal theory will imply that knowledge cannot be present in the actual circumstances. Thus, there may even be some reason to prefer (1) to (2). In any case, it is a mistake to think that a counterfactual theory containing (2) is superior to one containing (G1).

In conclusion, I submit that Lemke has given us no reason to think that the counterfactual causal theory can be resurrected. It either involves a counterfactual too close to being a necessary truth to impose any substantive requirements on belief basing, or it involves a counterfactual which makes the theory subject to simple extensions of Lehrer's original gypsy lawyer case.

574 Harrington Tower,
Texas A & M University,
College Station, Texas 77843-4237,
U.S.A.

© JONATHAN L. KVANVIG 1987

A FIRST-PERSON ASYMMETRY

By JAMES MONTMARQUET

TIME was when discussions of the first-person standpoint were almost entirely concerned with its special epistemic assets: with the privileges and apparent prerogatives of self-knowledge. More recently, attention has been drawn to the necessary limitations and deficiencies of such knowledge, e.g., with the 'epistemic blindspots' attaching to agent's predictions of their own decisions. (Cf. such essays of Roy Sorenson's as 'Uncaused Decisions and Pre-decisional Blindspots', *Philosophical Studies*, 45, 1984, pp. 51–6.) The present discussion falls within this latter 'tradition'.

Ordinarily, there is no great puzzle as to why I may know more than you about my own future actions, for I am likely to know more than you about my *intentions*. But suppose that both you and I know that I intend to go fishing Saturday. There is this argument that you may be in a *better* position than I to know whether I will go. You know, let us say, that when I have formed like intentions under like circumstances, only about half the time have I ended up making good on them; so you may reasonably conclude that, on the strength of the available evidence, the likelihood of my going fishing is about one-half. Of course, the same evidence (and, let us suppose, no other evidence) may be available to me; for I, too, may certainly be aware that I have only made good on such intentions about half the time. The question, though, is whether I can treat my *own* intentions in the sort of detached, purely 'objective' way that others can. Can I both intend to do x and treat *this very intention* as inductive evidence that I have, say, only about a 1 in 2 chance of doing x?

Consider what seems an analogous phenomenon governing belief. Suppose that an athlete, Jon McEnroe say, were to reason as follows.

> I feel entirely confident of beating Lendl. But wait, whenever I've been that confident before, I've beaten him about 80% of the time; so I should really be about 80% sure of winning. No, wait again, whenever I've been about *that* confident, I've beaten him about half the time; so I should be about 50% sure of winning. No, when I've been that confident in the past, I've beaten him only about 1 time in 10 . . .

If he continues in this vein, it is obvious that poor Jon's confidence (and consequent probability of winning, if the above relationship between these remains intact) will spiral downward until he reaches a point of stability: a point at which his degree of confidence matches what he takes to be the actual probability of his winning, given that degree of confidence. What seems less obvious here, however, is what McEnroe *should* believe in this situation. To be sure, from a standpoint of *practical* rationality, the answer to this is fairly clear: he should be as confident as he possibly can; for that will maximize, assuming that these conditional probabilities are correct, his chances of winning. But, from a standpoint of *epistemic* rationality, this is anything but clear. Even if he could get himself to be completely confident of victory, given that he believes (and justifiably, I will suppose) that with this degree of confidence he would stand only an 80% chance of victory, it is hardly clear that he *should* believe that he has anything more than an 80% chance of victory – but, of course, an analogous argument may be given as to why he is not entitled to *that* degree of confidence either, and so on. Is the answer, then, that McEnroe is only entitled to believe, say, that he has no serious chance of winning, if that is the point of stability – or is there some trick here?

If there is a trick, I think it is this. Supposing that McEnroe were 100% confident of beating Lendl (and were blissfully unaware of the above matrix), there is no question that *we* would be entitled to use that matrix in order to arrive at a justified belief that he has an 80% chance of victory. But consider what McEnroe is doing when he uses the matrix himself. Supposedly, he is seeking the epistemically correct level of confidence to have, given his present level of confidence. But, notice, unless we suppose that the time of this 'present level of confidence' and the time for which he is seeking the 'correct level of confidence' are distinct, this is incoherent. That is, if he is trying to ascertain what degree of confidence he should *now* have, given that he is now (say) 80% confident, this involves an incoherency. For suppose that he arrives at, and *believes*, the result given by the above matrix – then he will be committed to precisely this argument:

(i) I am now 80% certain that I will win.
(ii) If I believe that I have an 80% chance of winning, my actual chances are 60%; therefore:
(iii) I have a 60% chance of winning.

To accept both the premies and conclusion of this argument, evidently McEnroe would have to have inconsistent beliefs: he would have to believe both that he had an 80% and only a 60% chance of winning.

We may now return to the previous point about time parameters. Suppose that McEnroe is 100% confident at a given time *t*. Will there be a problem with this argument?

(i) I believed at *t* that I had a 100% chance of winning.
(ii) If I believe at *t* that I have a 100% chance of winning, my actual chances are 80%; therefore:
(iii) I have an 80% chance of winning.

Whereas no inconsistency lurks here (in that he can certainly *now* accept both the premises and conclusion of this argument), still, its second premise is questionable. Presumably the idea of the matrix was only that if McEnroe *goes into the match* with a given degree of confidence, he has a correlative chance of winning. Hence, it is to be presumed that merely believing 'at *t*' that he has a 100% chance of winning is not grounds enough for a reliable induction from previous occasions in which he has entered matches with this degree of confidence – certainly not when *t* is a time at which he is in process of revising his beliefs in order to tell how confident he should be going into the match.

On closer inspection, then, what seemed a paradox – that McEnroe should believe that he has no serious chance to win – turns out to be avoidable. Instead, what he should believe is that his chances are those dictated by the available evidence, *excluding* the above matrix. Suppose, then, that this value is 80%. Notice, then, that *we* as observers may know something which he does not, indeed cannot, namely: that his actual chances of winning, given *all* of the available evidence, are 60%. So we may well be in a better position than he to know whether he will win – even if all of us are apprised of exactly the same facts and principles of reasoning.

Return now to the case of knowledge of one's own intentions. Here a related argument will show that inductive reliance on one's own intentions is incoherent. Briefly, the argument is this. Even if intending to do *x* does not imply being *entirely* confident one will actually do *x*, certainly it will imply that one has *some* amount of conviction (however slight) that one will do the intended act. Now this degree of confidence will certainly be relevant evidence concerning the likelihood that the intended act will be successfully performed. But can the agent himself, as opposed to others, make use of it? The previous argument will have shown that he may not be able to. Indeed, it will have shown something perhaps more subtle, namely, that he can make use of his own subjective degree of conviction *only if he doesn't need to* – i.e., only if the probability of his winning, given all relevant factors including this degree of confidence, equals the probability of his winning, given all relevant factors except that one. (Note: if this equality does *not* obtain, then we have a McEnroe type situation; if it does, then obviously there is no need to make use of this additional information.)

We see, then, that persons can be barred – in some sense, necessarily barred – from the kind of knowledge of their own future actions that others may possess, utilizing the same evidence that agents themselves will have. A decade and a half ago, in a well known piece, Keith Gunderson spoke of the self as 'adept at slipping through the nomological net of physical explanations that philosophers have been able to imagine science as someday bestowing on them'. ('Asymmetries and Mind-Body Perplexities' in *Minnesota Studies in the Philosophy of Science*, vol. IV, University of Minnesota, 1970, p. 273.) Part of what we now see, I think, is that this same self may be captured in nets cast by others, but which (necessarily) it cannot cast itself.[1]

Tennessee State University,
3500 John A. Merritt Boulevard,
Nashville, Tennessee 37203,
U.S.A.

© JAMES MONTMARQUET

---

[1] This paper derives from a longer one written while attending an NEH sponsored Summer Seminar for College Teachers, H. N. Castaneda, Director, at Indiana University, 1986. I thank the Director and other participants for their helpful comments.

SENSE, REFERENCE, AND MEANING-INCOMMENSURABILITY

By STIG ALSTRUP RASMUSSEN

IN his otherwise stimulating book, *Representing and Intervening* (Cambridge U.P., 1983), Ian Hacking suggests that the realist in respect of theoretical entities needs to shun the Fregean account of linguistic meaning, if he is to stand a chance of rebutting a meaning-incommensurability argument, in the vein of Kuhn and Feyerabend, that poses a threat to his realism. Once he plumps for Putnam's account of meaning, in 'The Meaning of "Meaning"' (H. Putnam, *Mind, Language and Reality*, Cambridge U.P., 1975, pp. 215–71), the realist is however alleged to be in a position to block the argument. I shall argue that Hacking is wrong about the first claim: the argument is straightforwardly fallacious, on Fregean principles. Furthermore, Hacking is in any case unfair to the Fregean, who is in fact in a position to prevent the argument from taking off at all in much the same way this is effected within the Putnamian framework.

The meaning-incommensurability argument runs somewhat as follows:

(1) Semantical holism with respect to theoretical terms: 'The meaning of individual terms in (any) theory is given by their position within the structure of the theory' (p. 72; see also p. 76).

(2) Fregean assumption: the meaning of a word is its Fregean sense.

Hence,

(3) The sense of, say, 'The electron carries the minimum electrical charge' differs according as the sentence occurs within the context of Lorentz's theory or that of Bohr (p. 76; the example is discussed on pp. 82–84).

In the summary presentation of the argument on p. 76, Hacking stops here. However, (3) poses no *prima facie* threat to realism about theoretical entities; and on p. 73 he takes the further step necessary to generate such a threat:

(4) Those theories cannot be compared at all, '... because in principle they can never discuss the same issues'; or '... they are about different domains'.

Hacking locates the trouble in premiss (2) and suggests that adopting Putnam's account of meaning in terms of syntactic and semantic markers, stereotypes, and references/extensions will solve the problem (pp. 76–81).

The reason why the incommensurability argument is invalid on Fregean principles is this. As we know from Frege's celebrated Morning Star/Evening Star argument, words with different senses attached to them may share a common reference; whereas the move from (3) to (4) presupposes that they may not. True, in this instance, the terms involved sound and look the same; but the two occurrences of 'electron' are nonetheless, given the assumptions involved, from the Fregean perspective occurrences of different words, since the sense is different in the two cases. This may seem counter-intuitive and ill adapted in general to the fact that languages evolve over time. Let that be the objection, then: it is nothing to do with the incommensurability argument. (In fact, it is not clear that Putnam's account does much better on this score. If you take stereotypes seriously as part of word-meaning, the problem concerning the individuation of words remains. On the other hand, if you play up the role of references/extensions, as Hacking does, you have to invoke all the paraphernalia of initial dubbings and transmissions in order to avoid ending up with all co-referential words sharing the same meaning; and you have to do so in a manner such as to allow that 'electron' is really one word occurring in various theories ever since Johnstone Stoney, while 'the Morning Star' differs, as a term, from 'the Evening Star'. There are well-known difficulties, some of which are considered by Hacking, with obtaining the 'right' results.)

So, the Fregean would disallow the transition from (3) to (4). He would however have to admit that (1) through (3) will support the conclusion that 'electron' *may have* different senses within different theories; and also that scientists may have no immediate way of telling whether or not their respective pet theories refer to the same entities: they must take the trouble to study each other's theories. This, however, is no more than what one would expect anyhow; and the position is not significantly altered if we go Putnamian.

Furthermore, it should be noted that although we cannot, without (2), arrive at exactly (3), (1) *alone* allows the derivation of (3′):

(3′) The meaning of 'The electron carries the minimum electrical charge' differs according as the sentence occurs within Lorentz's theory or that of Bohr (this is in fact how Hacking formulates the lemma on p. 73).

If so, however, *any* theory of meaning that countenances the principle that meaning determines reference leads from (1) through (3) to the weakened version of (4) just adverted to; and Putnam's theory fits the bill truistically, since references/extensions form part of his quadruples constituting the meanings of terms. It is clear, therefore, that the pivot of the incommensurability argument is not the Fregean assumption, but, rather, the premiss that semantical holism holds in respect of a language comprising theoretical terms.

In accordance with this, the reason why Putnam's account, or something like it, blocks the derivation of (3) is precisely that his account offers a way of eschewing semantical holism. Theoretical entities are not directly observable, but they may be interpreted, in the light of theoretical assumptions that may not derive from the actual theory of the entities in question (cf. Hacking's chapter on observation, pp. 167-85), to leave 'traces' that are observable; and this is sufficient for fixing references – possibly by means of some set of original stereotypes that are not holistically bound up with the current theory of the entities.

The question then is whether or not the Fregean is committed to semantical holism. Curiously, Hacking admits that that doctrine is 'unFregeian' (p. 76); by which he presumably means that it seems to have had little appeal to Frege himself. Still, one can see how Hacking may have come to regard semantical holism as a doctrine germane to Fregeanism: when considering the history of naming mesons, Hacking takes the fact that the term 'meson' came to mean 'whatever it is that satisfies Yukawa's conjecture' as amounting to the term having acquired a Fregean sense (p. 90). If the Fregean is depicted as holding the sort of crude description theory of term-sense implied here, the imputation to him of semantical holism does indeed seem fair enough. However, even on the assumption that the Fregean is committed to the description theory, reference-fixing descriptions do not have to draw upon the content of the current theory of whatever entities are in question; and, more importantly, the Fregean does not have to embrace the description theory, at least not in general. The Fregean may take the scientists' grasp of the sense of 'electron' to consist in his knowing a procedure for identifying an electron; or, since that may in some instances be too tall an order, in his possessing a recognitional skill that enables him to identify the 'traces' of theoretical entities produced by others in some standard way. Since scientists holding widely differing theories about electrons may nevertheless agree over how standardly to detect their presence, they need have no trouble at all working out that they disagree, and how. On this picture, two groups of scientists who associate different reference-fixing procedures, i.e. different senses, with a theoretical term may in fact be talking about different entities; but this does not argue incommensurability, since there is no general reason why they should in principle be debarred from finding out.

University of Copenhagen
Købmagergade 50
1150 Copenhagen K
Denmark

© Stig Alstrup Rasmussen 1987

FREGE ON INFINITE AXIOM-SYSTEMS

By R. R. Rockingham Gill

PROFESSOR Dummett ('Frege on the Consistency of Mathematical Theories' in M. Schirn (ed.): *Studies on Frege* (1976), vol. I, pp. 229–31) is mildly reproachful of Frege for combining a 'recognition of the importance of infinite axiom-systems' with the view that 'the only way of proving the consistency of a theory is to provide a model of it'. As he says, one who thinks a theory consists of whatever is deducible from a set of propositions has another way available of proving that it is consistent (in the sense of there being no self-contradiction deducible from it), to wit, by showing that each finite subtheory of the theory is consistent. 'The point was... quite open to Frege to realise, and thus allow that the consistency of a theory may be proved without establishing that it has a model.'

Dummett thinks that Frege's argument in the fourth to sixth paragraphs of 'Logic in Mathematics' (G. Frege, *Posthumous Writings*, Blackwell 1979, pp. 202–3) 'provided him with a ground for considering infinite axiom-systems (systems containing axiom-schemes) as possibly sometimes being required', but also that Frege's recognition of them was 'purely hypothetical', since he, first, was a logicist, so that the argument referred to was merely *ad hominem*, and, second, 'had no qualms about higher-level quantification', so that 'axioms which in a first-order formulation would have to be presented as axiom-schemes he could express as single axioms'.

Now I question whether the argument referred to, whether *ad hominem* or not, is in fact to be interpreted as Dummett does, that is, as saying that: 'If... there is any form of inference peculiar to a particular mathematical theory, then we can incorporate it into the formalization of the theory by assuming, as an axiom, for every possible specific application of that rule of inference, the conditional whose antecedent is the conjunction of the premises of that application, and whose consequent is its conclusion.' What Frege *says* is that mathematical induction rests on a law: 'If this law can be proved, it will be included amongst the theorems of mathematics; if it cannot, it will be included amongst the axioms. If one draws inferences by mathematical induction, then one is actually making an application of this theorem or axiom; that is, this truth is taken as a premise of an inference.' It seems to me that 'this truth', 'this theorem or axiom', and 'this law' are here co-referential. What is their common reference? Frege gives something which could be taken to be the *schema*

$$((\theta 1 \wedge \forall n(\theta n \supset \theta n + 1)) \supset \forall n\, \theta n)$$

– but, of course, it can't be, for a schema does not express a thought; so he must have meant, rather, the *Grundsatz*

$$\forall \theta((\theta 1 \wedge \forall n\, (\theta n \supset \theta n + 1)) \supset \forall n\, \theta n)$$

Dummett presumably takes the three noun-phrases not to be co-referential, and supposes 'this truth' to refer to *the application of* the theorem/axiom – that is, to the instance of the schema which is actually used in the inference. That is certainly a possible reading of what Frege says, but it surely does not cohere with what he goes on to say: 'So likewise in other cases one can reduce a mode of inference that is peculiar to mathematics to a general law.... And from this law one can then draw consequences in accordance with general laws'.

So it appears that Frege's argument at this point gives no support to the idea that he 'recognized' infinite axiom-systems, and there was therefore no obstacle to his thinking that axiom-systems could be proved consistent in only one way.

One other piece of evidence which bears on the issue is what he wrote in § 5 of *The Foundations of Arithmetic* (Blackwell, 2nd edition 1953),

> We must distinguish numerical formulae, such as $2 + 3 = 5$, which deal with particular numbers, from general laws, which hold good for whole numbers. The former are held by some philosophers to be unprovable and immediately self-evident like axioms. Kant... hesitates to call them axioms because they are not general and because the number of them is infinite. Hankel justifiably calls this conception of infintely numerous unprovable primitive truths incongruous and paradoxical. The fact is that it conflicts with one of the requirements of reason, which must be able to embrace all first principles in a survey.

So at least at one time — admittedly thirty years earlier — Frege would have rejected the idea of an infinite system of unprovable truths. And an axiom was always for him a truth — if not unprovable, then unproved.

University of Wales,
St. David's College,
Lampeter, Dyfed.

© R. R. ROCKINGHAM GILL 1987

A PUZZLE CONCERNING DIVESTITURE

By STEVEN M. CAHN

SUPPOSE I hold 100 shares of stock in a company that has embarked on a policy I consider immoral. I therefore wish to divest myself of those 100 shares. In order for me to sell them, someone must buy them. But the buyer would be purchasing 100 shares of 'tainted' stock. And I would have abetted the buyer in this immoral course of action. Granted, the prospective buyer might not believe the stock 'tainted', but that consideration would be irrelevant to me, since I am convinced that, knowingly or unknowingly, the buyer would be doing what is immoral. Surely I should not take any steps that would assist or encourage the buyer in such deplorable conduct. Nor should I try to release myself from a moral predicament by entangling someone else. How then is principled divestiture possible?

Graduate School
City University of New York
33 West 42 Street
New York, NY 10036, U.S.A.

© Steven M. Cahn 1987

l. 47 No. 4 (New Series No. 216) October 1987

ANALYSIS

Edited by
CHRISTOPHER KIRWAN

CONTENTS

e truth about heaps	MICHAEL CLARK
tuitionism and sorites	STEPHEN P. SCHWARTZ
nse, *tense* and TENSE	GRAHAM PRIEST
eply to Spinks on temporal parts	HAROLD W. NOONAN
illiamson and the contingent *a priori*	GRAHAM OPPY
' again	ANTHONY APPIAH
quoque	ROGER TEICHMANN
eply to Noonan	E. J. LOWE
further reply to Noonan	B. J. GARRETT
real contrast	PETER LIPTON
ompound emphasis and causal relata	DOUGLAS EHRING
ozick's acceptance rule and the lottery paradox	R. P. LOU
n an epistemic paradox	DORIS OLIN
variation on Hollis' paradox	GEORGE REA
isualizing and imagining seeing	ALAN R. WHITE
rvival and importance	ANDREW A. BRENNAN
rfit's arguments against partially relativized theories of rationality	RICHARD M. GALE
ildren of a lesser God? A reply to Burke and Crisp	P. J. MCGRATH
eply to Wreen	JENNIFER HORNSBY
By" and "By"'	ALAN R. WHITE

BASIL BLACKWELL · OXFORD AND NEW YORK

ISSN 0003-2638

BLACKWELL

New books in the *Philosophical Theory Series*

General Editors: JOHN McDOWELL, PHILIP PETTIT and CRISPIN WRIGHT

Reality and Representation

DAVID PAPINEAU

David Papineau defends realism from a naturalistic perspective. He criticizes philosophers such as Hilary Putnam, Michael Dummett and Donald Davidson, arguing that representation is a natural, biological phenomenon which needs to be analysed in teleological terms.
264 pages, **£25.00** (0 631 15517 1)

Conditionals

FRANK JACKSON

A principal purpose of conversation, of writing and reading, is to transfer information and to effect modifications in the belief systems of hearers and readers. This book argues that an understanding of the various ways in which this can happen leads to a simple and satisfying theory of conditionals.
160 pages, **£19.50** (0 631 14621 0)

Abstract Objects

BOB HALE

Our familiar understanding of thought, reference and knowledge seems to leave no room for understanding abstract objects. In this original study, Bob Hale outlines the basis for a new Fregean Platonism in his defence of a new view of abstract objects.
224 pages, **£27.50** (0 631 14593 1)

New in the *Great Debates in Philosophy Series*

Agency and Necessity

ANTONY FLEW and GODFREY VESEY

In this debate, both authors agree about the reality of agency and personal responsibility but fundamentally disagree over how that reality is to be defended. The development of their dialogue is informed by the contributions of philosophers from Plato and Aristotle to Wittgenstein and Strawson.
192 pages, hardback **£25.00** (0 631 14539 7)
paperback **£7.95** (0 631 14541 9)

Basil Blackwell
108 Cowley Road, Oxford OX4 1JF
Suite 1503, 432 Park Avenue South, New York NY 10016

INQUIRY
An Interdisciplinary Journal of Philosophy

Editor: Alastair Hannay

Volume 30 (1987), Nos. 1-2
Foundations of Wildlife Protection Attitudes — Eugene C. Hargrove
You Can't Get Something for Nothing: Kierkegaard and Heidegger on How Not to Overcome Nihilism — Hubert L. Dreyfus and Jane Rubin
Meaning, Context, and Control: Convergent Trends and Controversial Issues in Current Social-scientific Research on Human Cognition and Communication — Ragnar Rommetveit
Concepts, Structures, and Meanings — G. R. Gillett
Overdue Analysis of Bourdieu's Theory of Practice — Theodore Richard Schatzki
Power and Formation: New Foundations for a Radical Concept of Power — David West
Two Dogmatists — Charles R. Pigden
Tarski - A Dilemma — Richard C. Jennings
Peacocke's Thoughts (Christopher Peacocke: *Thoughts: An Essay on Content*) — Richard E. Aquila

Subscription price:
Institutions: USD 60.-. Individuals: USD 31.-.

Published quarterly by Norwegian University Press, P.O. Box 2959 Tøyen, 0608 Oslo 6, Norway - Publications Expediting Inc., 200 Meacham Ave., Elmont, NY 11003, USA. Editorial address: Institute of Philosophy, P.O. Box 1024 Blindern, 0315 Oslo 3, Norway.

Teaching Philosophy

Tenth Anniversary Special

Teaching Philosophy, edited by Arnold Wilson of the University of Cincinnati, is celebrating ten years of publication by offering subscriptions to new individual subscribers at a special anniversary rate of only $10 per year. The regular subscription rate is $17 for philosophers and $38 for institutions. Orders must be pre-paid and received by November 30, 1987 to get this special low rate.

* * *

To help you discover the latest trends in the teaching and learning of philosophy, ***Teaching Philosophy*** publishes articles, discussions, reports, and reviews on these and related topics:

- theoretical issues in the teaching of philosophy
- experimental and interdisciplinary philosophy courses
- evaluation of teaching and assessment of learning in philosophy
- new books and audio-visual materials

* * *

Send orders to: *PHILOSOPHY DOCUMENTATION CENTER, Bowling Green State University, Bowling Green, OH 43403-0189.* ☎ *(419) 372-2419*

ANALYSIS 47.4 OCTOBER 1987

THE TRUTH ABOUT HEAPS

By MICHAEL CLARK

THERE is a simple way of explaining the Sorites Fallacy, or Fallacy of the Heap, which makes no explicit appeal to any quasi-intuitionist doctrine, makes no concession to scepticism, and poses no threat to the principles of mathematical induction and *modus ponens*. And it respects such obvious truths as that heaps are accumulations of grains, that baldness is typically the consequence of the loss of many hairs, and that adulthood is generally reached after the passage of many millions of seconds.

There is clearly no problem about the probative failure of the first of the following arguments.

(A) An account with a balance of −£100 is overdrawn.
For any (integral) n, if an account with a balance of n pence is overdrawn then so is an account with a balance of $n + 1$ pence.
Therefore an account with a balance of £0.01 is overdrawn.

(B) A single grain of sand does not constitute a heap.
For any (positive) n, if a pile of n grains of sand is not a heap then neither is a pile of $n + 1$ grains.
Therefore a pile of 100,000 grains is not a heap.

The second premiss of A is false, since a customer with a balance of −£0.01 is overdrawn whereas one with a balance of zero is not. By contrast the second argument is problematic because there is no critical number of grains which marks a sharp division between what is a heap and what is not. There is no range between overdrawn balances and those in credit analogous to the range of borderline cases of near-heaps. Doubtless there are hard cases in which it is indeterminate whether an account is overdrawn or not, but this will be due to indeterminacy over the balance, not to indeterminacy over whether an account with a given balance is overdrawn. On the other hand it seems quite natural to regard some accumulations of grains as mere approximations to heaps and their descriptions as heaps as only *roughly* true. As J. L. Austin pointed out in *How To Do Things with Words* (Oxford, 1962) there are many statements like 'France is octagonal' and 'Lord Raglan won the battle of Alma' which are only roughly correct. 'In real life, as opposed to the simple situations envisaged in logical theory, one cannot always answer in a simple manner whether [a statement] is true or false' (p. 142).

Some borderline piles are closer to being heaps than others: after a certain point, the attribution of the term 'heap' becomes more accurate as more grains are added, until there are so many that it becomes unequivocally true. Even the addition of a single grain brings the accumulation slightly closer to being a heap and makes the attribution a little more accurate, though the difference is so tiny that it will rarely concern us. Admittedly, an overdrawn account will get closer and closer to being in credit as it rises, yet it remains strictly true that it is overdrawn until it rises to zero.

Of course, the size and configuration of the grains as well as their number are relevant in determining whether we have a heap. Similarly, the passage to adulthood is a matter of physical growth which is not guaranteed simply by the passing of time, and the status of a scalp depends not only on the number of hairs but their thickness and distribution. But it simplifies the discussion to ignore these complications; and they can surely be accommodated without affecting the points at issue.

In the case of the overdraft there is a value of n for which the induction hypothesis fails. If there were a maximum number of grains, k, such that every accumulation of at most k grains failed unequivocally to constitute a heap, we could deny that the second premiss of the heap argument was true without qualification, since the antecedent of the following conditional would be true but its consequent only roughly true (even though as close an approximation to the truth as would ordinarily make no difference).

(H) If a pile of k grains is not a heap, then neither is a pile of $k + 1$ grains.

But it is indeterminate what k is. And even if it were not, it would still remain puzzling that such a minute deviation from the strict truth should yield such a patently false conclusion.

However, there will be a range of values of n for near-heaps, and although the bounds of this range are indeterminate this is presumably not so for all included values. Similarly, the bounds of adolescence are indeterminate but there are plenty of unequivocal cases of individuals on the threshold of adulthood. Now let k in the conditional (H) displayed above be the number of grains in some clearly borderline heap. Then both the antecedent and the consequent of (H) will be approximate truths, but the consequent will be marginally further from the truth than the antecedent. It does not seem unreasonable in such a case to regard the whole conditional as falling short of strict truth, by however small a degree. In that case there are many values of n, not just one, for which the generalization fails to hold unequivocally. And many of those could actually be specified. In the light of these considerations it seems fair to pronounce the generalization false. At the very least it is not true, nor does it come very close to the truth.

It is even plainer when the heap argument takes its original form as a sorites involving many steps of *modus ponens*. This will include 99,999 conditionals of the same form as (H), quite a few of which deviate very slightly from the truth. These tiny errors propagate through the chain to yield a wholly erroneous conclusion, as surely as the continued accumulation of grains generates a heap and as surely as we grew from children into adults with the passage of hundreds of millions of seconds.

The Sorites Fallacy can therefore be seen as arising from too strict an adherence to two-valued logic.

University of Nottingham,
Nottingham NG7 2RD

© MICHAEL CLARK 1987

INTUITIONISM AND SORITES

By STEPHEN P. SCHWARTZ

THERE seems to be a certain amount of confusion with respect to Putnam's proposal for dealing with sorites paradoxes by employing intuitionist logic. In a recent article Stephen Read and Crispin Wright attribute to Putnam the denial of

(1) $(\exists n)\ (\mathrm{BM}(n)\ \&\ {\sim}\mathrm{BM}(n+1))$

but Putnam denies that he denies (1).[1] Instead he claims to reject the inference of (1) from the denial of

(2) $(n)\ (\mathrm{BM}(n) \rightarrow \mathrm{BM}(n+1))$

and he claims also to accept

(3) ${\sim}{\sim}(\exists n)\ (\mathrm{BM}(n)\ \&\ {\sim}\mathrm{BM}(n+1)$.

The difficulty is in understanding how Putnam's intuitionist solution differs from a classical approach. Indeed, the substance of Read's and Wright's objection to Putnam's proposal is that it is 'classical' enough to allow sorites paradoxes to be generated. Even given Putnam's reply it is not entirely clear how his proposal is

---

[1] Stephen Read and Crispin Wright, 'Hairier than Putnam Thought', *Analysis* 45.1, January 1985, pp. 56–8. Hilary Putnam, 'A Quick Read is a Wrong Wright', *Analysis* 45.4, October 1985, p. 203. Putnam originally made his proposal in his article 'Vagueness and Alternative Logic', reprinted in *Realism and Reason*, Cambridge 1983, pp. 285–6.

'nonclassical' with respect to sorites. When one is confronted with an argument for a false conclusion – in this case (n) (a man with n hairs is bald) – then one has only two options: reject the validity of the argument or deny the premises of the argument. It would be expected that when an 'alternative logic' is proposed for dealing with a recalcitrant argument the argument's validity will be called into question. This, however, is not the case with Putnam's proposal.

The argument that Putnam considers in 'Vagueness and Alternative Logic' is as follows. The first premise is (2) above – $(n)(\mathrm{BM}(n) \rightarrow \mathrm{BM}(n+1))$ – and the second premise is the obvious truth $\mathrm{BM}(0)$. These two premises imply, using mathematical induction, the conclusion $(n)(\mathrm{BM}(n))$. The striking thing, however, in the present context about this argument is that it is both classically valid *and* intuitionistically valid. The validity of the argument is not in question under either approach. The basis for Putnam's rejection of the argument is not the logic of the argument but a premise of the argument. He denies the first premise. But this, obviously, is exactly what one would have to do under the classical approach (unless, of course, one accepted that everyone is bald). Given this, it is hard to see how the intuitionist[2] differs from the classicist. They are each confronted with an argument whose validity they accept but whose conclusion they reject. They both respond by rejecting the same premise. What, then, is the content here of 'alternative logic'?

One important difference that Putnam wants to emphasize is that the classicist, by denying (2), is committed to (1), which seems to be an obvious falsehood, whereas the intuitionist can duck this commitment. For the intuitionist the negation of (2) does not imply (1), whereas for the classicist it does. But now this may seem like splitting hairs. There may not seem to be much difference between having to assert an obvious falsehood and not being able to deny it. So if (1) is an obvious falsehood, the classicist and the intuitionist seem to be in almost equally hot water – one must assert it, the other can't deny it. This should not seem like much of a difference to one who is troubled by sorites paradoxes. If (1) is not an obvious falsehood, then the classicist may not be in much trouble by having to assert it and thus no 'alternative' would be needed. In either case, it is hard to see that much is gained by adopting the intuitionist approach here.

The significance of Putnam's intuitionist approach comes out more clearly, however, if we consider versions of the sorites that do not involve universal quantification or mathematical induction. It is well known that a sorites paradox can be given in a purely finitistic form using a chain of *modus ponens* or a long hypothetical syllogism. Consider the following version:

---

[2] By 'intuitionist' here I just mean someone who subscribes to Putnam's proposal to use intuitionist logic to solve the sorites paradox. I don't want to suggest (nor do I think would Putnam) that every intuitionist would embrace his proposal. Also it should be kept in mind that it is not entirely clear, as Putnam points out, what it means to employ intuitionist logic outside of pure mathematics.

(i) $BM(0)$
(ii) $BM(0) \rightarrow BM(1)$
$\vdots$
(n) $BM(10^{10} - 1) \rightarrow BM(10^{10})$
$\therefore$ $BM(10^{10})$.

This argument does not purport to prove that everyone is bald but it does almost as well, since it purports to prove the obvious falsehood that someone with 10^{10} hairs on his head is bald.

Again, this argument is both intuitionistically and classically valid, but now there is no premise, such as (2), that either the classicist or the intuitionist can happily deny. (2) is almost certainly false and Putnam claims to deny it, but it is hard to see how he can deny any of the premises of the present argument. Basically what we have done in the present version of the sorites argument is to replace (2) with a series of instantiations of it – enough instantiations to come up with an absurd conclusion. If the intuitionist denies any premise – say $BM(j - 1) \rightarrow BM(j)$[3] – then he is stuck with asserting

(4) $\sim\sim BM(j - 1) \;\&\; \sim BM(j)$.

But (4) seems just as arbitrary and unmotivated as what the classicist is forced to assert if he denies a premise – namely $BM(j - 1)$ & $\sim BM(j)$. It is just as arbitrary and unmotivated to make a distinction between not-not-bald and not-bald at j as it is to make a distinction between bald and not-bald at j. What argument could the intuitionist give to support his assertion of (4)? How could it be that one hair could make the difference between not-not-bald and not-bald? It is no more plausible that one hair could make the difference between not-not-bald and not-bald than it is that one hair could make the difference between not-bald and bald.

Still there is an important difference between the intuitionist and the classicist here. Both the intuitionist and the classicist must honour the commitment not to reject conclusions of valid arguments with true premises, so since both reject the conclusion, they must deny the premises, but it turns out that what it means to 'deny the premises' can be different in the two cases. In intuitionist logic $(\sim P \vee \sim Q)$ does not follow from $\sim(P \;\&\; Q)$. The intuitionist can deny a conjunction without being committed to denying any of the conjuncts. This means that the intuitionist can deny the premises by simply denying the conjunction of the premises. He is not further committed to denying any of the premises nor is he even committed to asserting that any of the premises are false. The most he is committed to is holding that not all the premises are true, but he doesn't have to say which ones are not true or even that it is determinate which ones are not true.[4]

---

[3] We need only consider conditional premises as $BM(0)$ is obviously true.

Here if anywhere is a significant difference between the intuitionist and the classicist. If the classicist denies a conjunction, then he is committed to asserting that at least one of the conjuncts is false. Thus when the classicist denies the conclusion of the hypothetical syllogism version of the sorites paradox, he is committed to denying one of the premises, in the face of all the difficulties involved in such a denial. Of course, he need not be able to say which one of the premises is false, but he must hold that one is false. Since the false premise will be one of the conditional premises, the classicist will have to hold that exactly one premise is false. For suppose $BM(j-1) \rightarrow BM(j)$ is a false premise; then $BM(j-1)$ is true and $BM(j)$ is false. We can assume that anybody with more hairs than a non-bald man is non-bald (we are also assuming that things like hair distribution and thickness are kept equal) and that anybody with fewer hairs than a bald man is bald. Thus each conditional premise after the false one will be true because it will have a false antecedent, and every conditional premise before the false one will be true because it will have a true consequent. The classicist is left with having to deny one and only one premise and thus is stuck with having no way to avoid an absolutely sharp line between bald and non-bald. The intuitionist seems to be able to avoid this unwanted consequence because he refuses to deny any premise (he just denies the conjunction of the premises) or even to agree that it is determinate which premises are not true (although he must assert that they are not all true).

But is the intuitionist being rational after all? If he rejects the conclusion of a valid argument and refuses to reject any of the premises, isn't he violating one of the fundamental rules of rationality? Not necessarily. One is committed to rejecting one of the premises of a valid argument with a false conclusion because of classical logic, not any fundamental rule of rationality. If the argument

$$\begin{array}{ll} & P_1 \\ & P_2 \\ & \vdots \\ & P_n \\ \therefore & C \end{array}$$

is valid, then the conditional sentence $(P_1 \,\&\, P_2 \,\&\ldots\&\, P_n) \rightarrow C$ is true (by the definition of 'valid argument'). By contraposition

---

[4] I would like to thank Professor Hilary Putnam for outlining, in correspondence, just how the intuitionist proposal would go here.

(which is accepted by both the intuitionist and classicist) we have $\sim C \rightarrow \sim(P_1 \ \& \ P_2 \ \& \ldots \& \ P_n)$. If one rejects the conclusion, i.e. asserts $\sim C$, then by *modus ponens* (also intuitionistically valid) we can derive $\sim(P_1 \ \& \ P_2 \ \& \ldots \& \ P_n)$. In other words, the notion of validity (together with the rules of inference recognized by both the intuitionist and classicist) commits us only to rejecting the conjunction of the premises of a valid argument with a false conclusion, but the notion of validity (together with those rules) does not commit us to rejecting one of the premises of a valid argument with a false conclusion. For that we need a further step that relies on the classical principle $\sim(P \ \& \ Q) \rightarrow (\sim P \vee \sim Q)$ which the intuitionist rejects. It may be that the intuitionist is wrong to reject that principle, but that is a different issue. The fact remains that the intuitionist has honoured his commitment to deny the premises of valid arguments with false conclusions. The crucial difference here, it turns out, between the intuitionist and the classicist is about what it means to deny the premises of an argument. The classicist is committed to denying one of the premises, the intuitionist is not. The intuitionist can simply deny the premises without denying any particular premise. The classicist cannot.

I think that the key to Putnam's proposal is to be found in the last sentence of 'Vagueness and Alternative Logic'. Putnam says:

> My proposal ... distinguishes between saying 'it is false that *all n* are such that — — —' and saying 'There exists an *n* such that not — — —' (p. 286).

In the case of the sorites we distinguish between saying 'It is false that all the premises are true' and saying 'There exists a false premise'. Of course, the classicist cannot make this distinction. This may be significant because this distinction may be crucial to the logic of vague terms. After all we know that the conjunction of the premises of the hypothetical syllogism is false but no conjunct seems to be deniable. Putnam's intuitionist proposal takes account of this situation in a way that is blocked for the classicist.[5]

Ithaca College,
Ithaca, NY 14850, U.S.A.

© STEPHEN P. SCHWARTZ 1987

---

[5] The majority of the work on this paper was done during an NEH Summer Seminar (summer 1986, Harvard University) led by Professor Hilary Putnam. I would like to thank Professor Putnam and the other members of the seminar, especially Lloyd Carr, Terence Horgan, Barry Loewer, and William Throop, with whom I had fruitful discussions of the topics in this paper. I would also like to thank Frederik Kaufman for helping me to clarify my thinking about these matters.

Tense, *Tense* and TENSE

By GRAHAM PRIEST

IN his article 'Tense's Tenseless Truth Conditions' (ANALYSIS 46.4, October 1986, pp. 167-72; subsequent page references are to this) Hugh Mellor replies to my objection to the argument of his book, *Real Time*, against the reality of tense. It is the purpose of this brief note to say why I think this reply fails.

Though I shall not attempt to make this note self-contained, let me set the background for the discussion. Mellor: observe that the truth conditions of all sentences (of a language such as English) can be given in a non-tensed metalanguage; tense is therefore not part of reality. Priest: the observation holds if and only if one assumes that truth is not tensed. If one assumes that truth is tensed the truth conditions of all sentences can be given in a tensed metalanguage. We therefore need an independent argument for the claim that truth is not tensed if the argument is to work.

In Mellor's reply I find essentially two rejoinders. Stating them requires some preliminary explanation. Mellor distinguishes between the grammatical tense of a verb form and pastness, presentness and futurity themselves. I, like he, will use upper cases for TENSE in the second sense, and italics for grammatical tense. Mellor points out, quite rightly, that TENSE and *tense* do not always align; that *tense* is, at best, a *prima facie* indicator of TENSE; and that in the debate about whether reality is tensed, it is TENSE that is important, not *tense*. According to Mellor, sentences, predicates etc. are TENSED in a derivative sense. A sentence (or more generally a truth bearer) is TENSED, if it 'explicitly or implicitly ascribes a TENSE to something' (p. 167); and a predicate is TENSED if 'it makes sentences containing it TENSED' (p. 168).

Mellor's first point is this: to give the truth conditions of non-*tensed* sentences in a *tensed* metalanguage, I used the verb form: *x eternally v's*, defined, stipulatively, as: *x has v'ed*, *x v's* or *x will v*. Mellor claims that eternal verbs are not, in fact, TENSED. Several points are relevant here. First, the verbs used in my definition are undoubtedly *tensed*. As Mellor observes, this does not automatically mean that they express (the relevant) TENSE, but in this case they obviously do so. Thus, eternal verbs are disjunctions of TENSED verbs; but whether they themselves are TENSED according to Mellor's definition is not so clear. The application of the definition is clear enough in the case of atomic predicates, but not in the case of molecular predicates. Do the predicates '*x* will not happen tomorrow' and '$2 + 2 = 4$ or *x* will happen tomorrow' 'explicitly or implicitly ascribe TENSE' to something? I don't know.

The only sensible way to settle the matter is to examine what the *point* of applying or withholding the epithet 'TENSED' is here. It is important to determine whether or not the sentences used in giving truth conditions are TENSED, since it is these which express possible facts. (On this Mellor and I agree.) Now what sorts of facts do compound sentences require? Facts just are whatever make true sentences true. And what is it in virtue of which compound sentences are true/false? As long as we stick to truth-functional compounds (which are all that are at issue here since eternal verbs are disjunctions of atomic verbs) the truth values of compound sentences are fixed once the truth values of their components are fixed. Thus compound sentences relate to the same facts as do their components. Compound sentences of TENSED sentences are therefore TENSED. In particular, then, eternal verbs are TENSED.

There is a possible, though perhaps baroque, defence here. This is to suppose that there are compound facts to which compound sentences relate; and moreover that these are not merely aggregates of the atomic facts of which they are composed (in whatever way compound facts are composed); but are independent enough to have emergent temporal properties (so that they are non-TENSED, though their parts are TENSED). This line of thought is not particularly appealing, employing as it does the reification characteristic of some of the more extreme forms of logical atomism; and I do not want to reheat these old chestnuts unless Mellor is inclined to do so.

There is another reason for not following this particular debate down labyrinthine paths. This is that the whole business is beside the point. Even if it were the case that eternal sentences express TENSELESS facts, it would remain the case that the metalanguage used in my construction refers to both TENSED and non-TENSED facts (TENSED facts being referred to by the truth conditions of *tensed* sentences). If this construction is right, therefore, TENSE is still an aspect of reality.

Conceivably an appeal could be made to Ockham's razor at this point: since the truth conditions in Mellor's original construction refer to only one kind of fact (non-TENSED ones), his approach is preferable. However, people who live in glass houses should not wield Ockham's razor. For a decent wielding would get rid of compound facts, if not reified facts altogether; and with them the charge that the *tensed* metalanguage is more complex since it appeals to two different kinds of fact.

It is therefore important for Mellor to have an independent case for the claim that truth is not TENSED (as I stressed in the conclusion of my original paper); and if this can be made, then my alternative construction can simply be ignored altogether. Such an argument is produced in the second part of Mellor's reply. To this I now turn.

Mellor's case that the truth predicate is not TENSED divides into two, depending on whether truth is predicated of sentence types or sentence tokens. Attribution of truth to sentence types is not TENSED since no sentence type has 'an A-series position, and calling it true never ascribes one to it' (p. 169). But it is not only objects that have positions in the A-series. It is also states of affairs. And one obvious TENSED interpretation of the claim 'sentence (type) *a* is true' is that the state of affairs specified by *a* has a certain A-series position, viz., now. Thus, 'The sun shines' *is* true iff the sun's shining *is realized* now. Similarly, 'The sun *shone* on the last day of 1986' *is* true iff the sun's having shone on the last day of 1986 *is realized* now. Under this interpretation the truth predicate is clearly TENSED, according to Mellor's definition. (Note also that the second example shows that *a* and '*a is* true', need not have the same TENSE, *pace* Mellor p. 169.)

Mellor, I assume, would object to this interpretation. Under it, '*x* is true' and '*x* is now true' would mean the same – or at least, have the same truth conditions. And Mellor argues in another part of the paper that '*x* is B' cannot mean '*x* is now B' on the ground that this leads to the 'obvious vicious regress' (p. 168). Indeed the identity does lead to a regress. But it is not vicious; and it does not show that the two things have different meanings. Similarly, the claim that A and ~~A mean the same entails a regress, but this does not imply that A and ~~A have different meanings.

Essentially the same argument can be used to show that '*a* is true' has a TENSED interpretation when *a* is a sentence token. Here, however, Mellor has another argument. He points out that if truth is TENSED then sentences may change their truth value over time. This is not, presumably, objectionable when truth is ascribed to sentence types. It is easy to see how a type (or even a long lasting token) can change its truth value over time. What however, in the case where the token is short lived? I say 'It is 1986'. This is true (as I write). But tomorrow it (the numerically identical utterance) will be false. I made plain in my original article that this was a consequence of the supposition that truth is TENSED. But Mellor finds it 'absurd to say ... [for example] ... that John's death posthumously verifies every premature announcement of it' (p. 170). But this is not absurd at all. What makes such an announcement premature is that it *was* false when it was made (though it *would be* true). Nothing that happens changes that. The utterance is verified in the sense of being made true (now), not in the sense of somehow showing the utterer to have been right then.

Mellor appeals to what he claims to be an analogous situation with spatial indexicals. Someone in Perth utters 'I am 10,000 miles from Cambridge'. This utterance is true, a situation that is not changed by the person travelling to Moscow. (Though it seems to me that one could systematically understand things in this way.) But someone who accepts the reality of tense will not be moved by this and similar examples. For they will reject the analogy. After all, it has always seemed to people that *tenses* (temporal indexicals) describe properties of reality in a way that spatial indexicals do not. There is no argument about the reality of hereness and thereness, for example. It is not, therefore, surprising that truth behaves with respect to time in a way different from that in which it behaves with respect to space. Thus this analogy just begs the question.

TENSED truth is (*is*) therefore not (yet) banished.

University of Western Australia,
Nedlands, Western Australia 6009

© GRAHAM PRIEST 1987

REPLY TO SPINKS ON TEMPORAL PARTS

By HAROLD W. NOONAN

IN my article 'A Note on Temporal Parts' (ANALYSIS 45.3, June 1985) I argued that *if* enduring objects have temporal parts, it is reasonable to say that some of the temporal parts of an enduring object of sort *S* (i.e. those long-lasting enough to exhibit the appropriate properties) are themselves objects of sort *S*. If persons have temporal parts, it is reasonable to say that some of these are persons; if tables have temporal parts, it is reasonable to say that some of these are tables; and so on.

Graham Spinks ('Noonan on Temporal Parts', ANALYSIS 46.4, October 1986) objects to this that temporal parts of enduring objects of some sort *S* cannot themselves be of sort *S* since the coming into and passing out of existence of a temporal part of an object (not matter how long-lived) of sort *S* is not a sufficient condition of the coming into or passing out of existence of an object of sort *S*. For example, the coming into or passing out of existence of a temporal part of a table is not a sufficient condition of the coming into or passing out of existence of *a table*.

This was one of the objections I had in mind when I said in the penultimate paragraph of my note that if my suggestion was not to be easily refutable it must come as part of a package. The other item in the package I mentioned there was the denial of the reducibility of restricted (sortal) quantification to unrestricted quantification, e.g. the denial of the equivalence of 'some table is ...' and 'something is a table and is ...' It should be evident that if this equivalence is denied Spinks objection is blocked: that some temporal part of a table came into existence at *t* will merely entail the truth of 'something *which is a table* came into existence at *t*', but not the truth of '*some table* came into existence at *t*'.

Of course, the denial of this equivalence is controversial (I said a little about it in my note and I have also discussed it in 'Relative Identity: A Reconsideration', ANALYSIS 46.1, January 1986, and

'Reply to Lowe', ANALYSIS 46.4, October 1986), and perhaps Spinks would not find it acceptable; but its availability is enough to show that his argument is not the knock-down refutation of my suggestion which he takes it to be.

One final point. At the beginning of his note Spinks writes 'Noonan's strategy is not to provide a positive argument for his claim, but rather to consider and discuss two commonly made objections'.

But I thought that I *had* provided a positive argument for my claim, at least with respect to the temporal parts of persons. Why is it *not* a positive argument for the claim that these qualify as persons that (as stated in the quotation from Lewis) they walk and talk and think and have beliefs and desires, hopes, feeling, wishes etc? What more needs to be said in support of the claim that some entity is a person?

University of Birmingham,
P.O. Box 363,
Birmingham B15 2TT

© HAROLD W. NOONAN 1987

WILLIAMSON AND THE CONTINGENT *A PRIORI*

By GRAHAM OPPY

IN ANALYSIS 46.3, June 1986, Timothy Williamson claims to have produced an instance of a 'deeply contingent *a priori* truth which in no way turns on indexicals' ([3], p. 113). However, I want to suggest that his argument does not establish this claim.

Williamson's argument is meant to establish that the deeply contingent sentence

(1) There is at least one believer

can be known *a priori* to be true in such a way that this knowledge does not depend upon one's knowing the truth of an indexical sentence such as

(2) I am a believer.

Williamson invites us to consider the following method for forming beliefs:

(M) Given a valid deduction from the premise that someone believes that *P* to the conclusion that *P*, believe that *P* and argues as follows:

(a) It is logically necessary that any belief one forms by use of (M) is true.
(b) Since (M) is an absolutely reliable method for forming true beliefs, a belief that *P* which has been formed by use of (M) is a case of knowledge that *P*.
(c) One can know *a priori* that (M) is absolutely reliable, so the knowledge which it yields is *a priori*.

Therefore

(d) If one comes to believe that *P* by use of (M), one knows *a priori* that *P*.

But

(e) If someone believes that (1), then (1) is true.

Therefore

(f) One knows (1) *a priori*, if one comes to believe it by use of (M).

Now, straight off, one is tempted to say that the statement of (M) in the imperative mood is merely a trick to avoid revealing the essential dependence on indexicality which is involved. For suppose that I am a person who has used (M) to obtain *a priori* knowledge of (1). Then it seems that there are two ways in which I could have argued:

(i) I could have started from the observation that if I were to believe that (1), then (1) would be true – i.e. I could apply (M) in the form:

(M′) If from the fact that I believe that *P* it follows that *P* is true, then I shall have a true belief if I believe that *P*.

But this would bring in the indexical element which Williamson claims to avoid.

(ii) I could have started from the observation that if, from the fact that someone were to believe that (1) it would follow that (1) is true, then someone who wanted to form a true belief should believe (1) – i.e. I could apply (M) in the form:

(M″) If from the fact that someone believes that *P* it follows that *P* is true, then anyone can form a true belief by believing that *P*.

But in this case, since (1) is an existentially quantified statement, and (M″)((1)/*P*) is a universally quantified statement, (1) is not entailed by (M″)((1)/*P*), and so I could not claim to know that (1) solely on the basis of (M″). Thus, even though (M″) is knowable *a* *priori*, the argument that (1) is knowable *a priori* is incomplete. Moreover, it seems clear that the argument could only be completed by my recognition that I am someone to whom (M″) is applicable – i.e. that I am someone who can form a true belief by believing that (1) – and so indexicality must be relied upon in this case too.

Now, I take it that Williamson would say that these considerations don't meet his argument – for what he has in mind is that someone is simply to form the belief that P whenever she has a valid deduction of the form specified in (M). If this is so then, although Williamson's argument is not perfectly transparent, his fundamental line of thought would seem to be something like this: let 'Bxp' mean 'x believes that p'. Then

APK $\quad Bxp \rightarrow p$

$\quad\quad\quad \therefore p$

is a pragmatically justified rule of inference – i.e. APK is analogous to the rule of

MP $\quad p \rightarrow q$

$\quad\quad\quad p$

$\quad\quad\quad \therefore q$

except that it is only truth-preserving *when used*. Now, in the case of MP we can allow that someone who detaches a conclusion by using this rule knows that the conclusion is true despite the fact that they can offer no justification for the claim that it is a reliable method of inference. For otherwise – as the Tortoise pointed out to Achilles – it seems that we will be led into an infinite regress of justification. So, in the normal case, all that is required in order for someone to know that the conclusion of an inference by *modus ponens* is true is that they should know that the premises are true. But then, why shouldn't APK have exactly the same status? In which case, if the premise is known *a priori*, then so is the conclusion.

One tempting line of objection to APK is this: a machine in a subject-less world which was programmed in accordance with MP would never produce false output from true input. However, if it were programmed in accordance with APK it could – and yet this should surely count as a use of APK. But then when I use APK, that in itself does not guarantee that the conclusion is true – and so it cannot be that I can *know* solely on these grounds that the conclusion is true – for surely it might be that I am merely a machine. Moreover, it won't do to object to this that I *do* know that I am not a machine – for in this context, this is tantamount to the claim that I know that I am a believer.

However, it might be said that this sort of objection is really a demand that someone who uses APK to obtain knowledge must know that it is reliable in her case; i.e. there is at least room for suspicion that it begs the important question. It is hard to see how one could show that a regress of justification in connection with APK is *vicious*, but there is nothing in the story about the machine which rules out the possibility that a higher-order pragmatic justification could be given. Thus, this line of objection is not *obviously* decisive.

Another possible line of objection to APK is this: Williamson claims that one could know (1) *a priori* on the basis of an inference by APK. But (1) is an existentially quantified sentence and, in many cases, if one claims to know 'There is something which is *F*', one's claim to know is undermined if one cannot answer the question 'Which thing?'. Now, consider a world in which there is only one subject. If that person cannot cite herself as the instance which makes (1) true, then surely she does not know that (1) is true. Hence a claim to knowledge of (1) on the basis of APK alone is undermined in the way suggested.

However, one can have classically valid non-constructive arguments for the truth of existential claims, turning on the law of excluded middle or other intuitionistically suspect rules of inference. And even though APK is logically independent of these rules of inference, the possibility remains that it may be fundamentally related to them. Moreover, it seems that it won't suffice to claim that an argument for (1) on the basis of APK is importantly different from a non-constructivist argument for the truth of an existentially quantified mathematical proposition (say), in that the one and only court of appeal for (1) is APK, whereas the mathematical proposition is supported by a proof, which in turn is supported by the methods and procedures of mathematics. For, even if the appropriate comparison is not between APK and *modus ponens*, but rather between APK and a mathematical proof, it might still be said that a demand for justification of a non-constructive mathematical proof and a demand for justification of an argument which uses APK have much the same status. That is, this line of objection also seems to be not obviously decisive.

A third possible line of objection to APK is this; consider again our world with only one subject. If that person cannot cite herself as an instance which makes (1) true, then what sense can we make of the idea that she *understands* (1)? If she does not know that she herself is a believer, then can she be said to have the concept of a believer? But then, what sense are we to make of the claim that our subject knows that (1), given that she does not understand it? Surely a necessary condition for the attribution of knowledge of a proposition to a subject — as opposed to attribution of knowledge that a certain sentence expresses a true proposition — is that the subject should be able to understand the proposition!

However, it might be said that a subject with limited powers of self-awareness may have concepts which she is not able to ascribe to herself provided that she is disposed to ascribe them correctly to others. That is, perhaps our subject's understanding of the concept of a believer would be manifested if there were other subjects of which she could predicate it. In which case, the suggested separation of knowledge and understanding has not occurred – which, in turn, means that this objection is not decisive.

Now, certainly, there is an air of paradox about this last reply. For, on the one hand, it is the subject herself who makes the sentence (1) true. And yet, on the other hand, her understanding of (1) is supposed to have *no* connection with the application of predicates to her own case. But, appearances notwithstanding, it remains to be shown that a contradiction does ensue.

Thus far, then, we have seen that provided Williamson sticks to his guns, it may be very hard – or even impossible – to establish that one who uses APK to detach (1) does not *know*, or does not know *a priori*, that (1) is true. Perhaps one may feel that the responses offered on Williamson's behalf are rather weak, but the burden of proof clearly remains with the one who wants to dispute the *bare* possibility that APK could be a fundamental rule of inference.

So what is wrong with Williamson's account? Well, consider our subject with limited self-awareness again. Why won't she use APK to reach the contradictory conclusion that a subject-less world must have a subject – i.e. why won't she apply APK in her counterfactual reasoning about a world in which there are no subjects? After all, except for her argument via APK she has *no* reason to think that there are any subjects in her world. Answer: because the description of APK contains a hidden indexical element which ties it to *her* world. In effect, the specification that APK is pragmatically valid amounts to saying that a conclusion inferred by use of it is true in whichever world is actual relative to that inference. Thus, in order to say how APK is to be used, one has to introduce – covertly or overtly – a reference to the world of utterance.

To this suggestion it might be objected that, in the absence of actuality operators and other specifically indexical devices, it is simply taken for granted that one is talking about the world of utterance – i.e. that, on the above account, *any* proposition is indexical. But this response misses the point. For it is a 'merely pragmatic' feature of these utterances that they are about the world of utterance. In order for this feature to have 'logical' significance, the indexicality has to be built in explicitly, in the way suggested.

Thus I suggest that an adequate representation of APK is something like this:

$$Bxp \rightarrow p$$

$$\therefore p\ (¢)$$

where (¢) is an operator, analogous to an actuality operator, which fixes the world in which p is to be evaluated as the world in which the utterance of p – and hence, of the whole argument – takes place.

Or, alternatively, one could read 'p(¢)' as 'p (in this world)', where the reference of 'this world' is determined by features of the context of utterance – in particular, by the world in which the context is located.

Of course, this way of looking at things suggests that the move from (2) to (1) as a sample of the contingent *a priori* trades one brand of indexicality for another. But, despite its initial oddity, this is in some ways a satisfying consequence – for (1) is then related to those sentences like

(3) If anything is red, then it is actually red

which satisfy the criterion which Evans proposes in [1] for the superficially contingent *a priori*. Of course, Evans' account of the contingent *a priori* will stand in need of some modification – for instance, in response to the objections which Forbes provides in [2] – but it seems to me that some account of the contingent *a priori* which turns on indexicality in the way that Evans' account of the superficially contingent *a priori* does will turn out to be correct.

University of Melbourne,
Parkville, Victoria 3052,
Australia

© GRAHAM OPPY 1987

REFERENCES

[1] G. Evans, 'Reference and Contingency', *Monist* 1979.
[2] G. Forbes, 'States of affairs, Bedeutung and the Contingent *a priori*', Typescript 1985.
[3] T. Williamson, 'The Contingent *a priori*: Has it Anything to do with Indexicals?', *Analysis* 46.3, June 1986.

'IF' AGAIN

By ANTHONY APPIAH

I

VIC DUDMAN has written more sense than anybody I know of about the taxonomy of English if-sentences. Anyone who wants to theorize sensibly about them should take a look at the many examples he has published and consider whether their theories can handle them. In 'Appiah on "if"' (ANALYSIS 47.2, March 1987) he criticizes my view in the light of some examples. In this response, I seek to explain away his major objections to my theory.

In my book on *Assertion and Conditionals* (Cambridge University Press 1985) I hadn't absorbed Dudman's contribution to the classification of if-sentences; and so I missed an opportunity to get right which particular English sentences my book was about. Let me begin by acknowledging and rectifying this error.

To begin with, I wish to rule out all examples in which 'if' can be paraphrased as 'whenever'; these are all cases which seem to involve some quantification over times, and the theory of my book was explicitly (p. 162) not meant to deal with them. Other if-sentences I shall call 'singular'. Dudman rightly distinguishes between singular if-sentences of the form 'If *A*, *C*', where *A* and *C* are expressions that themselves occur alone as English sentences and which occur in this context with the sense they have when they occur alone; and other English sentences of this form where the *A* or the *C* aren't themselves sentences or, if they are sentences, do not occur in senses that they can have when they occur as sentences alone. Where a clause occurs with a sense it can have as a sentence alone, I shall call it a 'declarative' occurrence; otherwise it occurs 'undeclaratively'. (This is a proposal that Dudman made earlier and has now complicated and improved: but I want to make points that I don't think depend on those complexities.)

Thus, we should distinguish between these sentences:

(2) If John comes, I will see him
(3) If John is coming, I will see him

While (3) has a reading that makes it equivalent to 'If it is true that John is coming, I will see him' – reflecting the fact that 'John is coming' occurs declaratively – (2) doesn't have a reading as 'If it is true that John comes, I will see him' – reflecting the fact that 'John comes' occurs *un*declaratively.

Dudman says that the information communicated by sentences like (3) is 'hypothetical': (roughly) because in them the antecedent occurs declaratively. The information communicated in other if-sentences he calls 'conditional'. In this paper I shall follow this usage, using 'hypothetical' and 'conditional' to refer to sentences in uses where they carry this information. (Dudman uses the term 'conditional' to apply to a narrower class of interpretations of singular non-hypothetical if-sentences; that doesn't matter here.)

It would be nice, having noticed this important distinction – a distinction of tremendous importance, as Dudman has shown, for the behaviour of tense in if-sentences – if we could follow Dudman in supposing that one semantic account applies to hypotheticals and others to various classes of conditional. I am not yet sure that this is right. But in this paper I shall apply my proposal only to hypotheticals.

Provided we deal sensibly with the anaphoric devices associated with pronouns and tense, any hypothetical will have a consequent that occurs declaratively. Thus: 'I will see him'. My theory, then, is supposed to apply to hypotheticals.

Now the analysis can run exactly as I said in the book. 'If *A*, *C*' is assertible iff p(*C*/*A*) is high. That is my view. I call this Adams' Hypothesis (AH, in brief), in honour of Ernest Adams who turned it into a very substantial theory. The claim is that speakers are entitled to assert such a sentence, where their conditional subjective probability for *C* on *A* is high. On this view (as I argued in Chapter 11 of my book) it is roughly true that 'If *A*, *C*' is assertible iff coming to have minimal evidence that *A* would put you in a position where you would assert *C*.

How should I deal with the objections Dudman makes? There are three. The first, to which I will turn in a moment, is that my theory makes the wrong prediction about the very case I originally considered. The second is that the sentence 'If I walk on my stairs, I will not fall through them' isn't one to which my theory applies. It was certainly meant to apply to this sentence. Dudman has persuaded me to give up AH for conditionals. This objection, then, I concede.

The third objection, which comes towards the end of the paper, is that Pascal's 'If we must not act save on a certainty, we ought not to act upon religion, for it is not certain' and three other sentences don't fit with the probabilistic analysis. I agree about Pascal's sentence: but the probabilistic analysis was only meant for cases where antecedent and consequent are truth-valued. I do not wish to assign a truth value to either here. The generalized version of the theory that I mention in the second part of the paper covers this case quite nicely, however. As for the other three cases, I don't see the difficulty in applying the probabilistic analysis. Two of them – his (10) and (11) – communicate messages in ways that are, in my view, essentially a matter of Gricean implicature. (12) raises no obvious problems at all.

Still, it may be worth saying, apropos of Dudman's general scepticism about the notion of assertibility, that it is intended to allow precisely for the distinction between the total message a speaker intends to get across on a particular occasion, on the one hand, and that core of the message, the grasping of which is the minimal mark of competence in the language, on the other. Take Dudman's (10):

> If it really is a law of nature that people resemble the pets they keep, then what is waiting at home for the Archdeacon is a cute little toy poodle.

My claim is that this is assertible iff the probability of the consequent given the antecedent is high. If I believed that the Archdeacon resembled a cute little toy poodle, then that conditional probability, here, would be high; provided, as we have every reason to believe, I think that whether there is such a law is independent of whether the Archdeacon resembles a cute little toy poodle. That is why this sentence is a good way of representing myself as believing that the Archdeacon resembles a cute little toy poodle. In normal circumstances (say, outside science fiction) it will be clear that the relevant resemblances are rather abstract. But figurative language, here as elsewhere, is to be understood in ways that depend, first, on our understanding of its literal uses.

That leaves the first objection.

II

Dudman condenses the heart of my original example nicely into three sentences.

> Suppose a plurality of urns containing little coloured balls, among them urn *B* containing 999,999 white balls and one black one. And suppose that, knowing this, we choose a ball *z* from an unknown urn, possibly *B*, with our eyes shut. Then, says Anthony Appiah ... we ought to assent when someone says
>
> (1) If *z* is from urn *B*, *z* is not black.

He then seeks to turn it into a counter-example. 'Now I am very sorry, but the hard fact is that we *not* assent to (1). Not in Australia, anyway, where I have rehearsed Appiah's scenario to literally hundreds of my fellow-citizens, mostly students.'

I constructed this as a hard case for my theory and, for the record, I now think that it confirms the old saw that hard cases make bad law. For it is misleading for a reason that has nothing to do with hypotheticals. That reason is that it shares an interesting feature with the so-called lottery paradox.

The point is this. In the lottery paradox, you'll recall, we have every reason to think, of each ticket, that it is extremely unlikely to win. (It is best to consider a lottery where the chance device that chooses the winner has not yet operated, so that uncertainty about which ticket will win is not just epistemic.) If any particular ticket's winning gets unlikely enough, according to some, we can say just plain:

(4) This ticket won't win.

Since we can say this of every ticket in a lottery where every ticket has the same chance, we seem to be committed to assenting both to a sentence of the form of (4) for each ticket and to 'One ticket will win'; which is the negation of the conjunction of the sentences of the form of (4).

Now, just as many English-speakers — and, if Dudman is right, all Australians in particular — won't assent to (1), however swamped the black ball is by the white ones, so many English-speakers won't assent to (4), however unlikely it is that each ticket will win. Just like his Australians, who say 'You're ignoring the (possibility that it's a) black ball', they say things like, 'You're ignoring the possibility that the ticket will win'.

My view is that the reason this is so, is that the general rule for assertion is not (as according to the official theory of *Assertion and Conditionals*) that you should assert what you believe to be very probably true; but rather that you should assert what you think you know. (I hinted at these problems in footnote 9 of the last chapter; p. 248.) And I think the reason why we are loath to assent to (4) is that, though we have every reason to have a high degree of belief that it is true, we have no reason to think that we know that it is true. The reason we think that we don't know it, is that *we have just as good a reason to say this of the ticket that will, in fact, win as of every other*. In all the cases where (4) will turn out to be right, its being right is, in some sense, accidental.

The case that I offered, involving the urns, is likewise one in which there is a small definite probability that we shall turn out to be wrong. Suppose that, by a miracle, it is the black ball in my hand. *I would have had just as good a reason to say (1) in this case, as in all the cases where I got a white ball* just as I expected.

How does this let me off the hook? Consider the official theory of Chapter 11 of my book: it says that you should utter a sentence in the appropriate class only if coming to have minimal evidence warranting the assertion of the antecedent would lead you to have grounds for asserting the consequent; reasons, then, for believing you know it. (This proposal, as I argued, has AH as a consequence, in cases where antecedent and consequent have truth conditions.) Well, though in this case evidence that you had urn *B* would give you reason to have a high degree of belief that you were not holding a black ball, you have no reason to think that evidence that you had urn *B* would give you *knowledge* that the ball was not black. The reason you wouldn't know it, is that *you have just as good a reason to say that the ball won't be black, in the case where it is black, as you would have in the cases where it is white*. In all the cases where saying 'It's not a black ball' – on the basis of evidence warranting the assertion that the ball came from urn *B* – will turn out to be right, its being right will be, in some sense, accidental.

The point, then, is that I can explain reluctance to utter (1) in ways that have nothing much to do with the semantics of if-sentences: and, better still, explain it in a way that *depends* on my analysis. That the reaction to (1) is not to be explained as simply as Dudman supposes can be reinforced by considering another related case.

Suppose, instead of my original set-up, we have a set-up like this: there are just two balls behind a door, and which one appears is determined randomly. The chances that the random device will select ball *A* and that it will select ball *B* are both 0.5.

Ball *A* is white, and, so far as we know, nothing that is likely to happen will change this. Ball *B* is white, too, but if the muon-flux rises above a certain level, it will change to black as the result of a complex physical-chemical reaction. The chance of the muon-flux being that high inside the box is precisely calculable as 1 in a million.

Would Australians assent in this case, as they wait the five seconds for the door to open, to the sentence

(5) The ball will be white?

If not, is there any level of probability, less than 1, at which they would? And if not, why do Australians have so much to say about other empirical matters of fact, since quantum theory – which is the best theory we have – predicts that very many future contingencies – the sun's rising tomorrow, for example – have some finite probability of not occurring? It seems to me that in this case reasonable anglophones ought to assent to (5).

That granted, modify the case to give a hypothetical. Suppose now that whether the door opens at all is also independently randomly determined; and that the probability that the door will open at all is also 0.5. Would Dudman's students say, back to the machine, when the five seconds have elapsed:

S: If the door is open, a white ball is visible?

I trust they would.

What does this show? Dudman says

> in (1)'s case, I contend, the reason for confessing to the hypothesis that *z* came from urn *B* is so as to be able to use it as a premiss for inferring the consequent proposition that *z* is not black. Seeing that this conclusion follows only if there is no black ball in *B*, I take the protestations of my respondents, cited above, as confirming my account as well as overturning Appiah's.

A lot depends on what 'follows' means here. If 'follows' means that

(6) The ball is white

is something which a person with our information should come to believe, in the original set-up, once told that 'The urn is *B*'; then it seems to me false to say that it follows only if there is no black ball in *B*. (I have already explained why someone might not *say* (6): this is because they have exactly the evidence they would have had if it had been black, which rules out knowledge, while allowing a high degree of belief.) If it means that it follows deductively from a specification of the situation that the ball is white, only if we have the premise that there is no black ball, then that is certainly true. But then, in the new case I have just set up, we should also need the premise that there was no black ball in the box: and, until we are *sure* of the muon-flux, we do not have that premise in this case either. Still, in *this* case, I hypothesize, his students *will* assent to S, even though it doesn't follow deductively from the information they have that the ball is white.

So it seems to me that by 'follows' Dudman must mean something like this: being told that the urn is urn *B* (or *that the door is open*) would warrant assertion that the ball is white (or *that there is a white ball visible*). But that is just the theory of Chapter Eleven of my book. It issues in Adams' Hypothesis on reasonable assumptions about what is involved in accepting evidence for an antecedent; and it is inconsistent with Dudman's positive proposal.[1]

Cornell University,
218 Goldwin Smith Hall,
Ithaca, NY 14853, U.S.A.

© ANTHONY APPIAH 1987

---

[1] Like Dudman, I am grateful for what I have learned from our correspondence. It is nice to be read so carefully. I should like to add, also, that I have learned a good deal about the connections between knowledge and assertibility from Slote, M., 'Assertion and Belief', in *Papers on Language and Logic*, J. Dancy (ed.), Keele, 1979.

TU QUOQUE

By ROGER TEICHMANN

SOME theories lay down certain normative conditions, to be met by theories in general (as well, perhaps, as by sentences, propositions, or whatever). For example, some forms of verificationism give criteria of meaningfulness for sentences, and theories, in general; and some (probably naive) forms of relativism might be seen as laying down that any adequate theory should *not* claim to state the Absolute Truth, or something of that sort. A possible attack upon such a theory is to say that the theory itself does not meet the conditions it lays down. Thus it has been claimed that the verificationist theory of meaning must, by its own lights, be meaningless, since neither verifiable nor falsifiable; and it has been asked what relativism is doing if not claiming to state the Absolute Truth.

The principle behind such an attack can be stated:

(P) Any theory laying down normative conditions for theories in general should itself meet those conditions.

Presumably, the force of the 'should' is this: that if a theory did *not* meet its own conditions, it could not possibly have whatever it is which it claims makes for (or amounts to) acceptability in a theory – and this fact would suffice for its absolute unacceptability.

(P) lays down certain normative conditions for theories in general. If we ask of (P), 'Does it meet its own conditions?', we have Ryle's 'namely'-rider problem. The conditions it should meet are those of meeting its own conditions, namely those of meeting its own conditions, namely... It appears that (P) cannot meet its own conditions. Does this make it unacceptable by its own lights, and hence unacceptable absolutely?

Sensitivity to an object-theory/meta-theory distinction might induce us to forbid the attempt to apply (P) to itself. This would have to mean re-formulating (P) so that it applied only to theories of 'lower levels'. If this option is available to (P), however, why should it not be available to all those theories with which it concerns itself? (P)'s whole bite is taken away once we allow a theory's theory-conditions not to apply to itself. One would be hard put to it to convince an opponent of (P) that (P) was peculiarly entitled to take advantage of the object-theory/meta-theory distinction: for this 'entitlement' would seem to arise simply from the fact that (P) will be undermined unless it *does* take advantage of the distinction.

Another tack is to point out that, if (P) does not meet its own conditions, then neither does it *fail* to meet its own conditions. For if we try to state what conditions it fails to meet, we come back to the 'namely'-rider problem. At any rate, if it were all right to answer 'What conditions does (P) fail to meet?' with, 'Those of meeting its own conditions', it would *seem* to be all right to use the same answer in response to the question, 'What conditions does (P) meet?'. So if (P) is re-formulated to require that no theory should *fail* to meet its own conditions, does it then perhaps pass the test for acceptability?

The opponent of (P) will probably bring back the 'namely'-rider problem at this point by asking the question, 'What conditions does (P) not fail to meet?'. We can try to defuse this question by saying that what the re-formulated (P) requires is that *there should not be* any theory-conditions, stated by a theory, which that theory fails to meet. If a theory satisfies such a negative existential, the question '*What* conditions...?' may be inappropriate. However, such inappropriateness will usually stem from a theory's not laying down any theory-conditions. If it does lay down such conditions, we can usually ask, 'Does it fail to meet its own conditions?'. (P) does lay down theory-conditions (how else could it do its job?); but the proponent of (P) seems committed to the view that this question is one we cannot ask — or at any rate, cannot answer. For there is *at least* as good reason to answer affirmatively as negatively.

Clearly, following through a defence of (P) which avoids letting off those theories which (P) is designed to attack requires the adoption of various questionable measures; its proponent is not on such firm ground as he might imagine. His '*tu quoque*' may rebound.

Perhaps the sorts of attacks I have in mind (upon verificationism, etc.) are best not conducted by appealing to any principle like (P) at all. An alternative procedure would be simply to show, in each case, how the suspect normative conditions laid down by a theory, taken together with certain plausible assumptions (such as that the theory itself is neither verifiable nor falsifiable), yield a *reductio ad absurdum*. The guilty premise, it would then be claimed, was that which stated the normative conditions, which must therefore be false. Getting the argument into *reductio* form will tend to require that the suspect theory-conditions be quoted, thus:

(1) The following sentence, S, is true (and hence meaningful): 'Any sentence which is neither verifiable nor falsifiable is meaningless'.

(2) S is neither verifiable nor falsifiable.

Ergo

(3) S is both meaningful and meaningless.

The only general principle involved here would be that no theory should assert something which, taken together with relatively more plausible assumptions, yields a contradiction; an obvious enough principle.

Merton College,
Oxford OX1 4JD

© ROGER TEICHMANN 1987

REPLY TO NOONAN

By E. J. LOWE

IN his 'Reply to Lowe' (ANALYSIS 46.4, October 1986, pp. 218–21, responding to my 'Noonan on Naming and Predicating', ANALYSIS 46.3, June 1986, p. 159), Harold Noonan contends that the collection of planks constituting a certain ship on a given day *itself qualifies as a ship* — notwithstanding the fact (which Noonan recognizes) that this collection of planks is not *identical* with the ship which it constitutes at that time. This gives rise to an apparent paradox which Noonan takes some pains to dispel. My own view is that Noonan's efforts are wasted, since the contention giving rise to the apparent paradox is mistaken.

Noonan defends his contention on the grounds that, in his view (which coincides with R. M. Chisholm's), the collection of planks constituting a ship on a certain day *itself* satisfies the definition given in Webster's dictionary of a 'ship' as a 'structure used for transportation in water'. I am not myself inclined to set too much store by dictionary definitions, but that aside I think it is debatable whether Webster really does support Noonan in this matter. According to Webster, to qualify as a ship a thing must be a *structure*. But if we turn to the definitions of a 'structure' supplied by the same dictionary we find two that seem to be relevant: 'something constructed or built' and 'something made up of more or less interdependent elements or parts: something having a definite or fixed pattern of organization'. Now, is a collection of planks *something constructed or built*? Surely not: rather, it is something whose members may (as it happens) be used to construct or build something else (e.g. they may be used to build a *ship*). Again, is a collection of planks *something made up of more or less interdependent elements or parts*? Surely not: for the planks in a collection of planks are not parts or elements of which it is 'made up' (they just *belong* to it), nor are they to any degree 'interdependent' as far as their membership of the collection is concerned (each belongs to it quite independently of the others). Finally, is a collection of planks *something having a definite or fixed pattern of organization*? Once more, surely not: for the planks in any given collection of planks may be organized and reorganized in indefinitely many different ways, none of which needs to have any permanence.

But perhaps it will be contended in response to my last remark that at least during such time as the planks in a given collection of planks are indeed organized or *structured* in a certain way (such as the way required to make a ship), *the collection itself* is 'structured', or 'has structure', and hence *qualifies as a structure*: the fact that the collection may subsequently be 'structured' in indefinitely many other ways or even in no way at all is, it may be said, quite irrelevant. However, I do not find this objection conclusive, for two reasons. Firstly, I do not see why we should be compelled to accept that a collection of things *itself* 'has structure' when the things that are its members are structured or organized in a certain way. Secondly, even conceding the first point, I do not see why the mere fact that its members are organized or structured at a given time should qualify a collection as a *structure*, even only *at that time*. For on one plausible understanding of Webster's (second) definition of a 'structure', something could not qualify as a structure at one time while not so qualifying at another. On this understanding, Webster is saying (as one might expect of a definition) that to qualify as a structure a thing must *essentially* – not just accidentally or contingently – have a 'definite or fixed pattern of organization'. But since the members of a collection may at various times be completely *un*structured, it is clear that on this understanding a collection could not *ever* qualify as a *structure* even if it were conceded that it would indeed be 'structured' at such times as its members were. Altogether, then, it is far from evident to me that Webster offers Noonan the kind of support he supposes.

We do not, however, need to appeal to Webster's or any other dictionary in order to rebut Noonan's original contention that the collection of planks constituting a ship at a given time itself qualifies as a ship (even only at that time). All we need to do is to ask this: is it or is it not the case that for something to qualify as a *ship* it should be capable of surviving at least some change in its component parts – such as its planks – through renovation or repair? Clearly the answer must be that it is. *Of course* a ship may lose or gain a plank without thereby ceasing to be. By contrast, the very identity of a *collection of planks* is determined by the identity of the planks that are its members – a fact which Noonan himself implicitly recognizes – and consequently it does not make sense to speak of such a collection surviving the loss or gain of a plank. Hence no such collection can ever qualify as a ship.

It is perhaps worth making the following supplementary point in order to clarify a possible confusion. So far I have been discussing whether a *collection* of planks may qualify as a ship, since it was Noonan's contention that it could. But perhaps what Noonan really intended to claim, or would prefer to claim if the choice were put to him, is that something that we might call a *structured* collection of planks may qualify as a ship – where the difference between collections and structured collections is modelled on that between sets and ordered sets. A structured collection would be an entity whose identity-conditions were defined not only in terms of the identity of its members but also in terms of their structural relations – so that if x and y were structured collections containing the same members structured in different ways, then x and y would be non-identical (just as the ordered triples $<1, 3, 5>$ and $<5, 3, 1>$ are distinct despite having the same three natural numbers as members). Clearly, a structured collection – unlike a 'mere' collection, as we might call it – would be something *essentially* having structure (to the extent at least that *it* 'has structure' in virtue of its *members* being structurally related). It should be obvious, however, that my argument of the previous paragraph serves equally to demonstrate that a structured collection of planks cannot qualify as a ship, because just as with a 'mere' collection it makes no sense to speak of a structured collection of planks surviving the loss or gain of a plank.[1]

University of Durham,
50 Old Elvet,
Durham DH1 3HN

© E. J. LOWE 1987

---

[1] I am grateful to the Editor of *Analysis* for alerting me to a defect in an earlier version of this note.

A FURTHER REPLY TO NOONAN

By B. J. GARRETT

CONSIDER, again, the following three possible histories of the Ship of Theseus pictured below:

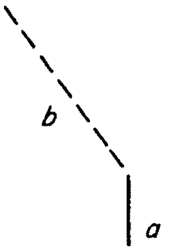

Figure 1

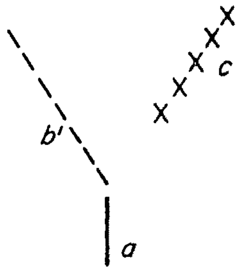

Figure 2

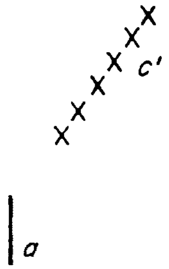

Figure 3

In all three situations a denotes the Ship of Theseus. In situation 1, b denotes the continuously repaired ship. In situation 2 (the situation described by Hobbes), b' denotes the continuously repaired ship and c denotes the plank-hoarder's ship, reconstituted from a's original planks. In situation 3, the ship reconstituted from a's original planks is denoted by c'. Given that we regard spatio-temporal continuity (under a sortal) as the most important criterion of artefact identity, outweighing the identity of original parts criterion, but acknowledge the fact that an artefact can survive dismantlement and reassembly, the most natural description of the above situations is as follows: in situation 1, $a = b$; in situation 2, $a = b'$ and $a \neq c$; and in situation 3, $a = c'$. This common-sense description coincides with that of the 'best candidate' theorist: the continuously repaired ship and the plank-hoarder's ship are both *candidates* for identity with the Ship of Theseus, though the former ship has the stronger claim.

Noonan has argued that the 'best candidate' theorist's description of situations 1–3 leads to absurdity (see [1] and [2]). If the 'best candidate' theorist respects the necessity and transitivity of identity – as, I now concede, he must – then he is:

> ... committed to saying that a situation differing from situation 2 in a mere Cambridge way with respect to what happens in the location of c – namely, by not containing the events there contained in the history of b' (a) after repair and replacement work begins – may be a situation from which c is absent ([2] pp. 208–9).

Noonan regards this consequence as self-evidently absurd. (He rightly points out, contrary to what he claimed in [1], that the 'best candidate' theorist's description of worlds 1–3 does *not* commit that theorist to the truth of the conditional that, but for the existence of *b'* in situation 2, *c* would not have existed *at all*, since situation 3 is *not* a situation in which *b'* (*a*) does not exist. However, Noonan notes that a 'best candidate' theorist *is* committed to the truth of such a conditional in the case of the division of persons, familiar from the literature on personal identity, in which there are two *equally good* candidates for identity with the original individual.)

Why is this consequence of the 'best candidate' theorist's description of situations 1–3, characterized in the above quotation, absurd? Noonan provides no argument – presumably he thinks that none is needed, nor can be forthcoming, since the absurdity in question is 'self-evident'. But it is not obvious to me that this consequence is absurd. Why should we not take the truth of the common-sense account of situations 1–3 as demonstrating, in effect, that the predicate 'constitutes the ship which is identical with *c*' does not denote any genuine property of the hunk of matter which constitutes *c*?[1] (This consequence makes clear that *c* is *not* a candidate for identity with Theseus's ship in the sense that there is a possible world in which *c* is Theseus's ship. Rather, it is the *hunk of matter* constituting *c* which is a candidate for being the hunk of matter which constitutes the Ship of Theseus; in situation 3 that hunk *does* constitute Theseus's ship.)

The following set of considerations ought, I believe, to give Noonan pause for thought. Many philosophers have held that not all predicates true of an object denote a genuine property of that object (e.g. the relational predicate 'being 50 miles south of a burning barn', to use an example of Shoemaker's). One way of attempting to give substance to this distinction has been to claim that genuine properties are those properties the possession of which by an object contribute to the *causal powers* of that object. It is the conceptual connection between the notions of (genuine) propertyhood and causality which is supposed to explain why a relational predicate such as 'being 50 miles south of a burning barn' does not denote any genuine property of an object of which that predicate is true (see [3]). *If* the 'causal powers' account is correct, then, by that criterion, *constituting the ship which is identical with c* surely ought not to count as a genuine property (would reference to such a property feature in an account of the causal potentialities of the hunk of matter which constitutes *c*?). Indeed, by the 'causal powers' criterion, the property *being identical with c* would plausibly not count as a genuine property of *c* – a consequence which has been embraced elsewhere, though on different grounds (e.g. on the grounds that any genuine property must be capable of multiple instantiations, see [4] pp. 29–31).

---

[1] An exactly parallel question can, of course, be pressed in defence of the 'best candidate' theorist's description of cases in which there are two *equally good* candidates for identity with an earlier individual.

Note that, strictly, it would be incorrect to claim that *constituting the ship which is identical with c* (or *being identical with c*) is a 'mere Cambridge' property. Geach's 'Cambridge Criterion' relates to changes, not properties (see [5] pp. 321–2). The concept of a mere Cambridge property can, of course, be defined in terms of the concept of a mere Cambridge change: a property *P* is a mere Cambridge property iff the loss or acquisition of *P* by some object *X* constitutes a mere Cambridge change in *X*. But, in our example, no object *changes* in respect of its possession of the property *constituting the ship which is identical with c* (and, obviously, no object can *change* in respect of its possession of the property *being identical with c*). One response would be to widen Geach's criterion as follows: a property *P* is mere Cambridge iff the loss or acquisition of *P* by some object *X* constitutes a mere Cambridge change in *X or* the possession by *X* of *P* in different possible worlds depends upon the occurrence of events (causally isolated from the events which constitute the history of *X*) in regions of space outwith the location of *X*. On such a modified criterion, *constituting the ship which is identical with c* will count as a mere Cambridge property.

Hence, unless Noonan *either* denies the validity of the distinction between genuine and non-genuine properties *or* offers an account of that distinction which is independently plausible and counts *constituting the ship which is identical with c* as a genuine property, it is open to the 'best candidate' theorist, in a way that is not question-begging (especially if the 'causal powers' account is correct), to argue that the consequences of his theory upon which Noonan has chosen to focus, far from being absurd, are actually in accord with our intuitive conception of what it is for a property to be genuine. On this account of the matter, there is no immediate need for the 'best candidate' theorist, if he is to avoid the charge of absurdity, to appeal to the thesis that identity is not 'what matters'; but it now presents itself as an interesting research project to attempt to chart the conceptual connections, particularly within the area of personal identity, between the theses that the identity relation is extrinsic, that predicates of the form 'constitutes the entity which is *X*' (or 'is identical with *X*') do not denote genuine properties, and that identity is not what matters.[2]

Lincoln College, Oxford OX1 3DR © B. J. GARRETT 1987

REFERENCES

[1] H. W. Noonan, 'Wiggins, artefact identity and "best candidate" theories', *Analysis* 45.1, 4–8.

---

[2] I am grateful to Peter Sandoe and Paul Snowdon for useful comments on an earlier draft of this reply.

[2] H. W. Noonan, 'Reply to Garrett', *Analysis* 46.4, 205–11.
[3] S. Shoemaker, 'Causality and Properties', in *Identity, Cause and Mind*, CUP, 1985, 206–33.
[4] A. J. Ayer, 'The Identity of Indiscernibles', in *Philosophical Essays*, London, 1959, 26–36.
[5] P. T. Geach, *Logic Matters*, Oxford, 1972.

A REAL CONTRAST

By PETER LIPTON

IN 'Explaining Contrastive Facts' (ANALYSIS 47.1, January 1987), David-Hillel Ruben considers the view that all explanations are really explanations of contrastive facts. On this view we never simply explain why *e* occurs, obtains, or is the case; instead we explain why *e* rather than *f*. Moreover, a good explanation of why *e* rather than *f* will not, in general, be a good explanation of why *e* rather than *g*. In Alan Garfinkel's illustration (*Forms of Explanation*, Yale University Press, 1981, pp. 21–2), the bank robber Willy Sutton explains why he robs banks by saying 'That's where the money is'. This explains why he robs banks rather than robbing other things, but it does not explain why he robs banks rather than leading an honest life. Ruben's position is that, while we may speak of contrastive facts, these are really just conjunctions of non-contrastive facts, and to explain contrastive facts is to give a certain type of explanation of the conjunctions. More precisely, Ruben holds that to explain why *e* rather than *f* we must satisfy two conditions. The first is that we explain *e* (whatever the explanation of a non-contrastive fact may require). The second is that we show that *e eclipses f*, where this seems to mean showing that *e*, along with other truths, entails that *f* does not obtain. To take Ruben's example, I explain why I was in London rather than Boston by first explaining why I was in London, then pointing out that since I was in London and no one can be in two places at the same time, I was not in Boston. I will argue that Ruben's account is unacceptable.

Strictly speaking, his account cannot be an analysis of the contrastive view he describes, since it is incompatible with that view; nor can it be used as an argument against the contrastive view, since it would then beg the question. The problem is that his account requires the non-contrastive *e* to be explained, but on the contrastive view non-contrastive facts are never explained. Still, it might be maintained that the contrastive view, as Ruben draws it, is extreme in this regard. It may seem that there are non-contrastive facts and that they cannot all be inexplicable. It may also seem natural to suppose that we can explain why *e* rather than not *e*, and that this is (as Ruben maintains) tantamount to explaining why *e simpliciter*. I do not myself know whether we can ever strictly explain non-contrastive facts, but let us continue on the assumption that we can.

At least three types of contrastive explanation provide counter-examples to Ruben's analysis. The first consists of contrasts that represent choices. If I am asked why I spent my vacation in London rather than in Boston, I may reply that the theatre is much better in London. Showing that going to London prevented me from going to Boston is beside the point. We cannot here explain why *e* was preferred to *f* by explaining *e* on its own and then showing that *e* eclipsed *f*. In the second type we have cases where *e* is a surprising result and *f* was expected. If I ask why the liquid turns blue rather than the red I had expected, you will not adequately answer my question by first explaining why it turned blue and then pointing out that if something is blue it cannot be red. The third type of counter-examples consists of cases where we want to know why two situations that seemed similar turn out differently. The familiar paresis/syphilis case provides a clear example. Even though few with untreated syphilis contract paresis, we can explain why Jones rather than Smith contracted paresis by pointing out that only Jones had untreated syphilis; for syphilis is causally necessary for paresis (cf. Bas van Fraassen, *The Scientific Image*, Oxford University Press, 1980, p. 128). Ruben's account fails here, since Jones's affliction hardly protects Smith.

Some cases of the contrastive explanation of choice, surprise and discrimination show that it is generally easier to explain why *e* rather than *f* than to explain why *e simpliciter*. That Jones had syphilis and Smith did not explains why Jones rather than Smith has paresis, but since so few with syphilis contract paresis, it does not explain *e simpliciter*, that Jones contracted paresis. This feature of contrastive explanations is one reason for the failure of Ruben's account. It also shows that explaining a contrast is not equivalent to explaining *e* and explaining not-*f*; nor to explaining the conjunction of *e* and not-*f*, since explaining the conjunction at least requires explaining the conjuncts.[1]

Williams College,
Williamstown,
Massachusetts 01267, U.S.A.

© PETER LIPTON 1987

---

[1] I am grateful to Tim Williamson for helpful suggestions.

COMPOUND EMPHASIS AND CAUSAL RELATA

By DOUGLAS EHRING

DONALD DAVIDSON has claimed (see [1]) that the sole causal relata are concrete events, spatial-temporal particulars which possess an indefinite number of features. Fred Dretske has denied this claim, arguing that it is not events but features of events which play the role of causal relata (see for example [2]). He bases his argument on the behaviour of causal contexts under emphasis. In this paper I argue that Dretske's argument runs into trouble in causal contexts involving higher degrees of emphasis.

Dretske begins his argument by claiming that a proposition will have different 'embodiments' depending upon different possible emphases. Hence the proposition 'Jones was shot at noon' has at least the following two 'propositional allomorphs':

(1) *Jones* was shot at noon.
(2) Jones was *shot* at noon.

Dretske argues that nominalizations of these propositional allomorphs will have different referents, in the form of different event features or event allomorphs. His argument for this thesis is that since substitution of nominalizations of differently emphasized allomorphs of the same proposition into the same causal context does not necessarily preserve the truth-value of the causal sentence, and since causal contexts are referentially transparent, it follows that such nominalizations differently refer. Further, causal contexts are sensitive to such shifts because the causal relation relates event aspects or event allomorphs and not events simpliciter. As applied to (1) and (2) this argument runs as follows. When we embed nominalizations of (1) and (2) in a causal context we may very well produce propositions of differing truth-values:

(3) The shooting of *Jones* at noon caused an uproar.
(4) The *shooting* of Jones at noon caused an uproar.

(3) might be true while (4) is false, or vice versa. Dretske argues that since causal contexts are referentially transparent, it is false that

(5) The shooting of *Jones* is identical to the *shooting* of Jones.

Again, according to Dretske what these expressions denote are event aspects rather than events, and causal contexts are sensitive to such shifts in expression because the sole causal relata are event aspects: '... causes (and, I might add, effects) are *facets* or *features* of events themselves' ([2] p. 375).

It can be shown that this argument will give rise to a dilemma: either Dretske's argument will generate a self-defeating regress or it will lead to the conclusion that causal contexts are not referentially transparent. In order to show that his reasoning gives rise to this dilemma, I will apply Dretske's argument to causal contexts involving what I will call compound emphasis.

Consider the possible referents of doubly emphasized expressions (where parentheses indicate double emphasis), such as 'the (*shooting*) *of Jones* at noon' and 'the *shooting of* (*Jones*) at noon'. Either it is the case that these expressions will each share their referents with these same expressions under some form of non-compound emphasis, e.g. 'the *shooting of Jones* at noon' or 'the *shooting* of Jones at noon', or we will have to recognize a different class of referents for these expressions under compound emphasis, i.e. higher-order features of events – in the case of double emphasis, aspects of aspects, e.g. not the heat-aspect of the fire, but the rate of increase in the heat of the fire.

If we take the referent of an expression like the 'the *shooting of* (*Jones*) at noon' to be a 'higher-order feature', in this case a second-order feature, an aspect of an event aspect, then Dretske's argument can be shown to give rise to a self-defeating regress. Specifically, if we grant that Dretske's argument shows, when applied to these nominalizations under non-compound emphasis, that events are not causal relata, then application to doubly emphasized nominalizations in the cause-position would show that event first-order aspects are not causal relata. But if this parity of Dretske's reasoning is granted, then we can reapply the same line to triply emphasized causal contexts to exclude second-order features, if there be such, from the class of causal relata. Indeed, for any level of emphasis, it will always be possible to find or construct causal sentences of that level which are amenable to further emphasis of varying patterns. And if doing as much at the level of unemphasized causal sentences shows that events are not causal relata, then by parity of reasoning we can eliminate features of any specified level by a similar application of Dretske's reasoning to causal sentences of that level of emphasis. On the assumption that to each higher level of emphasis there corresponds a higher level of features, we can thereby show that features of any specified level are not causal relata.

This objection, however, may depend upon an uncharitable reading of Dretske's argument, even granting the assumption that these expressions under compound emphasis refer to 'higher-order' features. The objection assumes that Dretske is arguing that if there is at least one unemphasized causal sentence S such that S may be emphasized in the cause-position (or effect-position) in more than one way, then events are not causal relata. On this reading, to generate the regress we only need to find *one* sentence, of a specified degree of emphasis, which can be subjected to varying patterns of further emphasis, in order to show that the entities corresponding to that degree of emphasis are not causal relata. But there is a more charitable interpretation of his argument which may succeed in bypassing this criticism. It is more likely that Dretske is arguing as follows. First, for any true unemphasized causal sentence, if it is grammatically possible to subject it to more than one pattern of emphasis in the cause-position (or effect-position), such that the result cannot be further emphasized, then its referring terms do not pick out events. Second, *every* unemphasized causal sentence may in fact be the subject of varying patterns of emphasis, but not every emphasized causal sentence may be further emphasized in varying patterns. Third, our causal ontology should be read off from the causal sentences which are not subject to any further emphasis. It will follow then, on the assumption that higher degrees of emphasis correspond to higher levels of features, that events are not causal relata, but that features of various levels are relata. On this reading of his argument it is not enough, as in the above criticism, to claim that for any degree of emphasis a sentence is constructable of that same degree of emphasis which may be the subject of further, varying patterns of emphasis. Dretske can respond to this objection by saying that our ontology is read off from the causal sentences, of whatever level, in which no further emphasis is grammatically possible, i.e. which are unambiguous in the relevant respect, and that this holds true even if there exist other causal sentences of that same level which are ambiguous in this sense. Finally, he could claim that there are no true unemphasized causal sentences in which emphasis is not possible, but that there are causal sentences of varying degrees of emphasis in which no further emphasis is possible.

This reply, however, even if successful against the charge of a self-defeating regress, opens Dretske up to another line of objection. On this version of the argument events must be admitted as causal relata since, contrary to what is claimed, we can find or construct *unemphasized* causal sentences in which no emphasis is possible. Consider the following example. Suppose I stipulate that I will refer to the shooting of Jones at noon as 'John' and the uproar as 'Fred', then the sentence 'John caused Fred' will be true. In turn, this latter sentence cannot be emphasized more than one way, in the cause- and effect-positions, such that the result is a sentence in which no further emphasis is possible. Or if this sentence seems contrived, or perhaps involves a misuse of proper names, consider the true sentence – 'this caused that', where I am using the first demonstrative to refer to the shooting of Jones at noon and the second demonstrative to refer to the uproar (I owe this example to Ray Elugardo). If we are to read off our ontology from the true causal sentences in which varying patterns of emphasis are not possible, then the latter sentence will commit us to events, presumably. Hence, even on this more charitable interpretation of Dretske's reasoning, it cannot be granted that he succeeds in ruling in features while excluding events as causal relata.

We may now return to the other possible reading of the referents of 'the (*shooting*) *of Jones* at noon' and 'the *shooting of* (*Jones*) at noon', the reading according to which these expressions will each share their referents with expressions under some form of non-compound emphasis such as 'the *shooting of Jones* at noon'.

On this rendering, additional emphasis marks no shift in reference, e.g.

(6) 'the (*shooting*) *of Jones* at noon', and
(7) 'the *shooting of* (*Jones*) at noon',

both refer to whatever

(8) 'the *shooting of Jones* at noon'

refers to (presumably the event aspect 'being a shooting of Jones'). However, on the assumption that the cause-position in causal statements is referentially transparent, it follows from the fact that

(9) The *shooting of* (*Jones*) at noon caused an uproar,

and

(10) the (*shooting*) *of Jones* at noon caused an uproar

may have different truth values that

(11) the (*shooting*) *of Jones* is not identical to the *shooting of* (*Jones*).

must be true. In other words, this assignment of referents can be maintained only if causal sentences are not referentially transparent, assuming (9) and (10) may differ in truth-value, an assumption to which Dretske would presumably agree. Without the assumption of transparency, however, Dretske's argument against events cannot get off the ground.

A proponent of Dretske's view might of course reply at this point that compound emphasis has no effect on truth-values in causal contexts, and, thereby, avoid much of this criticism. A defence along these lines, however, can be mimicked by the proponent of events at the level of non-compound emphasis. If we are willing to grant that compound emphasis has no impact on truth-values, a similar denial can be made about non-compound emphasis in causal contexts. In addition, it does seem to be the case that compound emphasis can make a difference in truth-values. Compare the following causal claims:

(12) The *shooting of* (*Jones*) at noon caused an uproar.
(13) The (*shooting*) *of Jones* at noon caused an uproar.

We can easily imagine a circumstance which makes (12) true but (13) false. Let us first suppose that the crowd which is upset by the incident is not moved by the fact that it occurred at noon. This accounts for the lack of emphasis in both (13) and (12) on 'at noon'. Secondly, the crowd is upset by the fact that the incident involved a shooting rather than some other kind of violence. We may imagine that while they have grown accustomed to other kinds of violence, shootings have been quite rare. Thirdly, what was even more upsetting than the fact that the violence involved a shooting was the fact that the object of the shooting was the much beloved Jones. The latter fact carried more causal weight in giving rise to the uproar than the former fact. Given this set of circumstances, although (12) is true, (13) is false. Or again we can imagine a circumstance which makes (12) false and (13) true, e.g. where Jones' involvement is causally important but less so than the fact that there was a shooting.

Our discussion shows that causal sentences under compound emphasis will require Dretske either to embrace a self-defeating regress, if he admits high-order features as referents of doubly emphasized expressions, or to abandon his assumption that causal contexts are transparent. In either case Dretske will not have shown that event features rather than events are the sole causal relata.

Southern Methodist University,
Dallas, Texas 75275, U.S.A.

© DOUGLAS EHRING 1987

REFERENCES

[1] Donald Davidson, 'Causal Relations', *Journal of Philosophy*, 64 (1967), 691-703.
[2] Fred Dretske, 'Referring to Events', *Midwest Studies in Philosophy*, 2 (1977), 369-78.

NOZICK'S ACCEPTANCE RULE AND THE LOTTERY PARADOX

By R. P. LOUI

IN *Philosophical Explanations* (Harvard, 1981), Nozick constructs the following acceptance rule:

Accept h on evidence e if
$\text{Prob}(e, h) \geqslant 0.95$;
$\text{Prob}(e, \text{not-}h) \leqslant 0.05$;
e; and
$\text{Prob}(h) > \text{Prob}(e, \text{not-}h)$. (p. 257)

In a note, he writes:

> I have not investigated whether the principle of inference we propose, all epicycles included, is open to any version of the lottery paradox. (p. 696)

He proposes the rule *caveat emptor*, with paragraphs-ful of apologies. It is almost criminal to pursue the possibility that the rule is open to lottery-like difficulties.

However, for the record, it is vulnerable to something like the lottery paradox. This is interesting because the recent rule of Good and McMichael ('A Pragmatic Modification of Explicativity for the Acceptance of Hypotheses', *Philosophy of Science 51*, 1984) is also vulnerable.

This means that of the proposed rules that threshold, i.e. that depend critically on one real-valued number, there is only one that avoids lottery-like inconsistencies. This is the Hintikka and Hilpinen rule. This rule avoids lottery-like inconsistencies by design, but avoids them in a manner that is specific to their system of confirmation; i.e. their solution would not work for other authors' systems (see Henry Kyburg's 'Conjunctivitis', *Epistemology and Inference*, Minnesota, 1984).

Interest in the lottery paradox has been revived in recent philosophical works of fairly wide appeal, including Stalnaker's *Inquiry* (MIT Press, 1985) and Harman's *Change in View* (MIT Press, 1986). Lottery-like difficulties have traditionally been used to cast doubt on principles of inductive inference. Recently, some members of the artificial intelligence (AI) community have also expressed interest in the lottery paradox. The paradox has surfaced in AI discussions of knowledge representation and in 'non-monotonic logic', which is AI's version of *ceteris paribus* reasoning.

Recall that the lottery situation is one in which there is a set of sentences $\{p_i\}$ such that each sentence is acceptable, while together they entail a contradiction; i.e. $\{p_i\} \vdash$ '$x \neq x$'. If an acceptance rule is open to lottery-like inconsistencies, then either (1) the rule cannot be used to determine rational beliefs, or (2) rational beliefs cannot be closed under the familiar entailment relation.

Here's a lottery-like example for Nozick's rule.

Let there be a large lottery; there are at least 401 tickets in the lottery. Let x be the probability, for any given ticket, that it will be the winning ticket. $0 < x < 1/400$. h_1 is the hypothesis 'the winning ticket number, n, is $1 \leqslant n \leqslant 20$'. h_2 is the hypothesis 'the winning ticket number, n, is $21 \leqslant n \leqslant 40$'. The probability of each of these is $20x$. e is the proposition 'the winning ticket number, n, is $1 \leqslant n \leqslant 40$, but $n \neq 10$, and $n \neq 30$'. The probability of e is $38x$. Figure 1 shows the relevant joint probabilities.

Fact 1. h_1 is acceptable on e.

(1) $\text{Prob}(e, h_1) = 19x/20x = 0.95 \geqslant 0.95$.

(2) $\text{Prob}(e, \text{not-}h_1) = 19x/(1-20x) \leqslant 0.05$. This is literally $400x \leqslant 1$, which is guaranteed by the size of the lottery, above.

(3) $\text{Prob}(e, \text{not-}h_1) = 19x/(1-20x) < \text{Prob}(h_1) = 20x$. This is literally $380x^2 + 20x - 1 < 0$, which is true for our small x.

	e	not-e	
h_1	$19x$	x	$20x$
h_2	$19x$	x	$20x$
not-h_1 & not-h_2	0	$1 - 40x$	$1 - 40x$
	$38x$	$1 - 38x$	1

Figure 1. Joint probabilities

Fact 2. h_2 is acceptable on e. The calculations are the same for h_2 as for h_1.

Fact 3. $h_1, h_2 \vdash$ '$x \neq x$'. In other words, h_1 and h_2 are inconsistent, by inspection.

So if e is the evidence, then a lottery-like inconsistency arises.

Later, Nozick discusses the h_i 'which fall under not-h', and adds the requirement:

> there is no h_i which might be true if h were false, that is also unexcluded by previous applications of the principle (that is, not-h_i hasn't already been concluded via the principle). (p. 258)

The h_i are supposed to entail not-h, not merely be propositions that 'might be true if h were false'. (The latter interpretation is absurd: some proposition, g, might be logically unrelated to h, yet might be true if h were false; we do not thereby exclude consideration of g's acceptance).

Correctly interpreted, this requirement avoids lottery-like difficulties explicitly.

However, it does so by adding a diachronic condition on acceptance. We cannot know which among h_1 and h_2 will be accepted, since this depends on which the agent considers first. Diachronic solutions to the lottery paradox have been known to the acceptance milieux for a while, but acceptance theorists generally have not considered them satisfactory. Nozick is an exception. His departure from synchronicity is reiterated in his 'Simplicity as Fallout' (*How Many Questions? Essays in Honor of Sidney Morgenbesser*, Hackett, 1983).

So what is interesting about Nozick's programme for acceptance is not the 'epicycles' in his acceptance rule. Rather, perhaps, it is the possibility of his successfully carrying out a diachronic programme of acceptance.

Departments of Computer Science and Philosophy,
University of Rochester,
Rochester, NY 14627, U.S.A.

© R. P. LOUI 1987

ON AN EPISTEMIC PARADOX

By DORIS OLIN

A THINKS of a number and whispers it to *C*; *B* does the same. *C* then announces *T*: '*A* and *B* have each chosen a different positive integer and neither can determine whose is greater'. According to Martin Hollis, this situation gives rise to a paradoxical train of thought (ANALYSIS 44.4, October 1984). In a previous paper, I argued that Hollis' paradox falls to the same solution as that of the surprise examination, and is therefore just another version of this familiar paradox (ANALYSIS 46.1, January 1986). Recently, however, Hollis has objected to my solution and presented two new variations to show that the paradox is 'wider than' that of the surprise examination (ANALYSIS 46.4, October 1986).

The source of the disagreement, I believe, is that Hollis has failed to recognize (and I had failed to make clear) the extent to which this is an *epistemic* paradox – the extent to which it rests on epistemic assumptions. Consider. For Hollis, the paradox arises as follows. Suppose *C*'s announcement, *T*, is true. *T* implies that neither party chose 1; for if *A* had chosen 1, then he would realize that *B*'s number was greater. But recognizing this, it also follows that no one chose 2; if *A* had chosen 2, he would know that *B*'s number was greater. Similarly for any positive integer. This seems to be a *reductio* of *T* – and yet, clearly, *T* can be true.

But the fact is that the paradoxical reasoning cannot even begin to get off the ground with just the assumption of *T*. Suppose *just* that *T* is true, and *drop* the fact that it has been announced. There is now no reason whatsoever to think that if *A* had chosen 1, he would know *anything* about *B* – whether *B* chose a number, whether he chose an integer, and so on. What is necessary, in order to get the train of thought going, is to be able to claim that *A* and *B* are each *justified in believing T*. This is why *C*'s announcement is crucial to the problem; presumably, his testimony would give *A* and *B* good reason to accept *T*. (The danger of confusing the supposition that *T* is true, with the supposition that *A* and *B* are aware of the truth of *T*, is avoided if we imagine ourselves, rather than *A* or *B*, working through the reasoning of the paradox.)

The announcement is thus crucial to the development of the problem. So let us consider now what obtains when *C* has announced *T* to *A* and *B*. Certainly, it seems plausible to say:

> If *A* had chosen 1, then he would be entitled to believe that *B*'s number was greater.

This first step of the argument goes through only if *A* would, under the circumstances, be justified in believing *T*. The core of my solution, however, consists in arguing that *if A* had chosen 1, he would *not* be entitled to accept *T*. The argument, which can be appreciated *independently* of the desire to escape the paradox, can be put thus. If *A* had picked 1, and were justified in believing *T*, then he would also be justified in believing:

> *B*'s number is greater than mine, and I am not now justified in believing that *B*'s number is greater than mine.

However, it is impossible to be justified in believing any proposition of the form '*p* and I am not now justified in believing *p*'. Hence, if *A* had picked 1, he would not be entitled to accept *T*. (For a fuller account, and an indication of how to handle variations, see my article 'The Prediction Paradox Resolved', *Philosophical Studies* 44, September 1983.)

My analysis consists, then, in showing that the paradoxical argument requires an epistemic assumption which, I argue, cannot be true. Notice that *A* cannot justifiably believe *T* *only* if he has chosen 1; and there is nothing to prevent someone other than *A* and *B* from believing it. It is also worth noting how the analysis applies to the simplest form of Hollis' paradox, which consists in announcing to *A*: '*B* has just chosen the number *x* and you, *A*, are not justified in believing this'. Although this assertion can be true, and can be justifiably believed by others, it cannot be justifiably believed by *A*. These results may be odd or surprising, but they are not paradoxical.

I leave it to the reader to verify that the analysis applies equally to Hollis' two further variations.

Glendon College, York University,
2275 Bayview Avenue,
Toronto, Canada M4N 3M6

© Doris Olin 1987

A VARIATION ON HOLLIS'S PARADOX

By GEORGE REA

MARTIN HOLLIS'S 'A Paradoxical Train of Thought' appeared in ANALYSIS 44.4, October 1984. Solutions were offered by Doris Olin and Michael Kinghan in ANALYSIS 46.1, January 1986. Hollis dismissed those solutions in ANALYSIS 46.4, October 1986. I shall argue that Hollis's reply is unsatisfactory, but that a variation on the paradox which evades the solutions of Olin and Kinghan may be constructed. I shall then offer a solution to that variation.

My description of the background to the paradox does not differ significantly from Hollis's recent modification. Both *A* and *B* know that *C*'s hat contains tickets numbered 1 to 300, and that no number is duplicated. *A* picks out a number and shows it privately to *C*. *A* does not replace his number. *B* then picks a number and shows it private to *C*. *C* then makes the following claim.

(D) Neither of you can determine whose number is the greater.

Suppose *A* picked 157. *A* would know that he at least could not determine whose number is the greater, since for all *A* knows *B* could have picked either a higher or a lower number, and that *B* may well be in the same sort of situation. D may well be true. But then *A* reflects that D implies that neither has picked 1 (or 300). If anyone had picked 1 then from his background knowledge (about the conditions of the number picking and his own number) he could determine that he had picked the lower number. Thus far *A*'s reasoning is quite right, but then he thinks that if a pick of every number less than *n* is incompatible with *C*'s claim of D being true, then so is a pick of *n*. *A* thinks that if someone picked *n* and realized that D implies that no one picked a number less than *n*, and *C* made a true claim of D, then the person who picked *n* could determine that he had picked the lower number (which is contrary to D). But in this case *A* is mistaken. Suppose *B* picked 2. *B* can *infer* from his background knowledge *and D* that he has picked the lowest number, but as Olin and Kinghan quite rightly pointed out this does not imply that *B* can *determine* that he has picked the lowest number. In order to determine this *B* would need to *know* that D is true. Moreover as Olin pointed out, the proposition that *B* has picked 2 and knows that D is true is absurd. *B* has picked 2, so he cannot know that D is true. In order to illustrate this, consider the following simplified inferences.

INFERENCE ONE

Common background knowledge	1.	*A* and *B* both know that they both know that they have each picked a different whole number

C's claim, D	2. Neither *A* nor *B* can determine whose is the greater number	
B's background knowledge	3. *B* knows that he has picked 2	
1, 2	4. *A* has not picked 1	1, 2
1, 2, 3	5. *B* has picked the lower number	1, 3, 4

INFERENCE TWO

Common background knowledge	1. *A* and *B* both know that they both know that they have each picked a different whole number	
Common knowledge of D	2a. *A* and *B* both know that neither of them can determine whose number is the greater	
B's background knowledge	3. *B* knows that he has picked 2	
1, 2a	4a. *B* can determine that *A* has not picked 1	1, 2a
1, 2a, 3	5a. *B* can determine that he has picked the lower number	1, 3, 4a

5a and 2a are incompatible. 5a follows from 1, 2a and 3. So 1, 2a and 3 must be inconsistent. 1 and 3 are true, so 2a cannot be true. However, there is no such inconsistency in 1, 2 and 3. 5 and 2 can both be true so long as 2a is false, which it is. Thus, contrary to Hollis's reply to Olin, the fact that *B* can make inference one from D does not imply that D is paradoxical. Rather D is true and *B* is not justified in believing it.

This criticism of Hollis's paradox rests on a definition of 'determine' such that iff *x* can infer P from some set of premises all of which *x* knows to be true, then *x* can determine P. If this is what Hollis means by 'determine', and the alternative phrase 'work out', then the objections of Olin and Kinghan which I have illustrated above will apply. However, there is an alternative variation on Hollis's paradox which evades this line of reply. *C* may alternatively make a claim about what *A* and *B* cannot infer, such that inference does not imply knowledge. Hollis's recent further examples seem to leave open this alternative interpretation, since he uses terms therein which do not so obviously imply knowledge, viz. 'able to decide' and 'able to infer'. However, such a variation on Hollis's paradox would require an understanding of what premises are allowed in the inference cited, since we would no longer be concerned with premises which are simply known. The main premise besides background knowledge which Hollis seems to have in mind is *C*'s claim. Suppose, instead of D, *C* makes the following claim.

(P) Neither of you can infer whose number is the greater from your background knowledge and this proposition.

Although superficially P may appear to be a simple statement, we find a serious complication of interpretation when we try to understand what premises beyond background knowledge are allowed in the inference cited. Consider the following.

One proposition which P seems to entail is P1.

> (P1) Neither of you can infer whose number is the greater from your background knowledge.

P1 will be true exactly when D is, i.e. iff no one picked 1 or 300. But P is supposed to mean more than this. P says that the premises of the cited inference may include not only background knowledge but also P, and hence anything entailed by P. So if P entails P1 then P entails P2.

> (P2) P1, and neither of you can infer whose number is the greater from your background knowledge and P1.

(P2 will be true iff both picked a number greater than 2 and less than 299.) For the same reason it follows that if P entails P*n* then P entails P*n* + 1.

> (P*n*) P*n* — 1, and neither of you can infer whose number is the greater from your background knowledge and P*n* — 1.

(P*n* will be true iff both picked a number greater than *n* and less than or equal to 300—*n*). Since P entails P1, this implies that P entails P2 and P3 and P4 and so on. In this case P cannot be true for any pair of numbers picked. For instance suppose *B* picked the lower number, say 34. P34 is false, since *B* can infer whose number is the greater from his background knowledge and P33. P entails P34, so P is not true. Since P cannot be true P is not paradoxical.

The same general solution may be applied to Hollis's further examples concerning Sheikh Inbed's distribution of goods: so I shall not go into detail. The solution depends on how the inferential term, e.g. 'able to decide' is interpreted. If this means 'able to determine' as in D then that line of solution will apply. If on the other hand it means 'able to infer from this proposition' then impossibility will become evident as with P.

8, Highfield Road,
Cleator Moor, Cumbria CA25 5NT

© GEORGE REA 1987

VISUALIZING AND IMAGINING SEEING

By Alan R. White

SEVERAL recent, philosophically perceptive, writers have alleged that to visualize or picture something, for example a face, a tree or a mountain – or, what they have wrongly supposed is the same, to imagine it – is to imagine either seeing it or oneself seeing it or that one sees it. Thus, Ryle says that '"having a mental picture of Helvellyn" or "having Helvellyn before the mind's eye" (or what he also calls "imagining", "imaging" or "picturing" Helvellyn) is actually a special case of imagining, namely imagining that we see Helvellyn in front of our noses' ([3] p. 256; cp. pp. 248, 251, 256, 266, 273); Vendler says 'imagine the battlefield from above or the statue sideways' is elliptical for 'imagine seeing the battlefield from above or the statue sideways' ([4] p. 47; cp. pp. 51, 59, 60); and Peacocke says that to imagine a tree implies at least 'to imagine the sort of experience one has when one sees a tree' and also to 'Suppose-imagine ... that the experience is a perception of a tree' ([1] pp. 23, 28; [6] seems to equate imagining a scene with imagining visual experiences).

Plausibility is added to this view, first, by the fact that the experience of visualizing may be like that of seeing. For Aristotle and Hobbes imagining was the having of a weak or decaying sensation. Ryle argued that 'a person picturing his nursery is like that person seeing his nursery' ([3] p. 248) and another philosopher has said that 'in visualizing I am as it were seeing' ([7] p. 35). Psychologists frequently suggest that the difference between visualizing and seeing is only one of degree and that both have the same physiological basis ([2] *passim*). Secondly, we report what we visualize or picture, just as we often report what we experience in memory, illusions, hallucination or dreams, in terms of 'seeing' in quotes; for example 'I still see the look of horror on her face', 'I just can't see him without a beard'.

Though these writers either move about indifferently between or explicitly equate imagining, for example, a face or a tree and visualizing or picturing it – a move which I think is mistaken – I shall concentrate on the idea of visualizing and argue that even in this more plausible version the common view is mistaken.

Two preliminary remarks are in order. First, it is at least clear that the use of 'imagine' in the analysis of visualizing a face or a tree as imagining seeing, oneself seeing or that one sees the face or the tree cannot itself be visualizing. For one reason, though one can, perhaps, visualize oneself seeing – or at least looking at – a face or a tree, to visualize a face or a tree is not to visualize or picture oneself seeing it, since one need not, and usually does not, come into the picture at all. For a second reason, to visualize a face or a tree cannot imply to visualize seeing, an experience of seeing or that one sees the face or the tree. A face or a tree is something one can picture, either in one's mind or on paper, but neither seeing, an experience of seeing or that one sees a face or a tree is picturable either in one's mind or on paper.

The second preliminary remark is that whatever the relation between visualizing something and imagining seeing, oneself seeing, or that one sees it, the latter clearly does not imply the former. One can say to someone in a surprised encounter 'Just imagine seeing you here', but not 'Just visualize you here', much less 'Just visualize seeing you here'. To say 'I never imagined seeing you here (or that I would see you here)' is not to say anything about visualizing. One can also imagine oneself seeing people dying of hunger and doing nothing about it without visualizing either oneself or them. And one can imagine that one sees one's friend in the distance without visualizing him in the distance. It is significant that though imagining hearing, smelling, touching (or oneself hearing, etc., or that one hears, etc.) is as common as imagining seeing (oneself seeing or that one sees), that 'hear', 'smell' and 'touch' in quotes are as common as 'see', and that imagining sounds, smells and feelings is as common as imagining sights, yet there is no 'auralize', 'olfactorize' or 'tangiblize' co-ordinate with 'visualize'.

Certainly, to visualize something is, as the word suggests, to take account in some way of its visual features; of the way it looks, did look, would or could look in certain conditions. It may even be to imagine what I would see if I were to look or to remember what I did see when I did look. But this is not to imagine seeing, myself seeing or, much less, that I see it. Imaginary sights and sounds are sights and sounds which I imagine to exist, not necessarily sights and sounds which I imagine seeing or hearing or that I see or hear. To say 'You can imagine the colour, smell, state of the body after it had been three months in the water' is not to say 'You can imagine seeing, smelling (yourself seeing or smelling, or that you see or smell) the state of the body . . .'. Analogously, to remember, describe, picture or paint, the visual features of something is to remember, describe, picture or paint, what it looks like, but it is not to remember, describe, picture or paint, seeing, my seeing, or that I see it.

When describing what one visualizes, or imagines, one often speaks of the characteristics of what one 'sees' ('hears', etc.), but not of the characteristics of one's 'seeing' ('hearing', etc.). 'This is how I see it' describes what is 'seen', not the 'seeing'. I can imagine (myself) suddenly, unexpectedly, delightfully or disappointingly, faintly, in a daze, seeing a tree, but not visualize or imagine a tree thus. I can imagine seeing a scene through the eyes of another, but I do not try to visualize anything either through my own eyes or through the eyes of another, though I can visualize it as it would look to this or that person from this or that angle. Vendler tries to avoid this sort of objection by arguing that, for example, 'Imagine seeing the Eiffel Tower through binoculars' – which, he admits, cannot be equivalent to 'Imagine the Eiffel Tower through binoculars' – really means 'Imagine looking at the Eiffel Tower through binoculars' ([4] p. 50). But this is not so. To imagine seeing something with the naked eye is not to imagine looking at it with the naked eye. To ask someone to imagine seeing something dreadful, shocking, beautiful or awe inspiring, may be to ask him to imagine having a kind of experience, but to ask him to imagine a dreadful, shocking, beautiful or awe inspiring scene itself is not. There may be many things I cannot imagine (myself) seeing and not reporting on or being overjoyed or disgusted at, but I can easily visualize them without imagining myself reporting, enjoying or being disgusted.

There is a difference, overlooked by Ryle ([3] p. 248), between a child who looking at her doll imagines that she sees a smile on its face and a child who thinking about her doll *in absentia* imagines, visualizes or pictures her smiling. One reason for Ryle's assimilation of imagining or visualizing something to imagining seeing or that one sees it may be a false analogy which he draws with recalling something. He says 'I recall only what I myself have seen, heard, done and felt, just as what I imagine is myself seeing, hearing, doing and noticing things' ([3] p. 273). But, first, though I may recall only what I myself have seen, etc., I do not in recalling what I have seen have to recall seeing or myself seeing it. Secondly, I can imagine others, as well as myself, seeing, hearing, doing and noticing things without in so imagining, imagining myself seeing them seeing, hearing, doing and noticing these.

To 'see' or 'picture' oneself or someone else as so and so, e.g. as a violinist or a vice-chancellor, as a satyr or a naiad, is to imagine oneself or another as these, not to imagine seeing oneself or another as these.

Peacocke suggests that his analysis of imagining or visualizing a tree as imagining the experience of seeing a tree is strengthened by its analogical analysis of imagining a pain in one's foot in terms of imagining the experience of feeling a pain in one's foot ([1] p. 31). But the important difference between imagining a tree and imagining a pain is that whereas there is a difference between a tree and seeing a tree and, therefore, a difference in imagining these two, there is no difference between a pain and feeling a pain and, therefore, no difference in imagining the two.

The analogy between imagination and memory, which has been pressed by almost every philosopher from Aristotle to Ryle, Vendler and Peacocke, is helpful here. When I try to visualize my friend without a beard, not because I am trying to imagine him without a beard but because I am trying to remember him without it, it is him without his beard that I am trying to remember, not (my) seeing him without it. It is his face, not my experience of his face, I am trying to recall. And when I finally say 'I can see it clearly now' or 'Its only a blank', it is my success or failure in recalling his face, not any experience of seeing his face, which I am marking. The fact that I describe my memories of what I have seen in such terms as 'I can see it now as clearly as if it were yesterday' suggests that I am remembering or recalling vividly those scenes, not that I am remembering or recalling my seeing of them.[1] I may be able to remember my friend's face without his beard, but not be able to remember seeing him without it, though I realize that if I can recall him without it, I must at some time have seen him without it. I may remember the experience of seeing him without it as being a shock or a pleasure, as being frequent or rare, but these are not characteristics of him without his beard. In trying to remember his face without a beard, I am trying to remember how it did look, not my seeing it. So in trying to imagine a bearded face without the beard, I am trying to imagine how it would look, as contrasted with how a remembered face did look, not my seeing either.

What the use of visual language, e.g. 'I can see it clearly now', 'I see him in my mind's eye', etc., shows is not that to visualize something is to imagine the experience of seeing or that one sees, but that the experience of visualizing seems like the experience of seeing. To suppose that visualizing a tree is imagining an experience like seeing a tree rather than having an experience like seeing a tree is to make a transposition analogous to that which Bertrand Russell made in moving from the assumption that in seeing the sun one has an image of the sun to the conclusion that in supposedly seeing the sun what one really sees is this image of the sun.

There is, I conclude, a clear difference between imagining or visualizing a tree which I imagine seeing, myself or someone else seeing, or that I see, and imagining or visualizing a tree which I do not imagine seeing, myself or someone else seeing, or that I see.

The University,
Hull

© ALAN R. WHITE 1987

REFERENCES

[1] C. A. B. Peacocke, 'Imagination, Experience and Possibility: A Berkeleian View Defended', in *Essays on Berkeley*, ed. J. Foster and H. Robinson, 1985.
[2] A. Richardson, *Mental Imagery*, 1969.
[3] G. Ryle, *The Concept of Mind*, 1949.
[4] Z. Vendler, *The Matter of Minds*, 1984.
[6] R. Wollheim, 'Imagination and Pictorial Understanding', *Proc. Aristotelian Soc.* Suppl. 60 (1986).
[7] Bernard Williams, *Problems of the Self*, 1973.

---

[1] In fact my memory image of the scene will not be the same as my original perception of it. I see myself with my friends in memory, but I did not see myself in the original group. Cp. Wittgenstein, *Zettel*, s. 650; Sartre, *L'Imaginaire*, II.5.2.

SURVIVAL AND IMPORTANCE

By ANDREW A. BRENNAN

IN a recent discussion paper (see [4]) Douglas Ehring criticizes Parfit's account of the psychological connectedness and continuity that is as good as survival on the grounds that it fails to satisfy a non-triviality requirement. In its general form, such a requirement insists that important relations (survival, identity, psychological connectedness or whatever) cannot depend for their holding on merely trivial circumstances. Since, according to Parfit, merely trivial differences are liable to make an impact on whether identity holds or not, identity cannot be of the importance that most of us suppose; in particular, when thinking about our own, and other people's, lives personal identity is not what matters (see [5], §§90, 91). I think that Ehring's criticisms are largely justified; indeed, they parallel ones that I make elsewhere (see [3], chapter 9.6, 9.7). What I will do in the present note is venture a diagnosis of why the problem described by Ehring arises in Parfit's treatment. If my diagnosis is correct, then it will help to explain my conviction, argued in a previous article, that personal survival is no more determinate a notion than personal identity (see [1]). The problem, it turns out, is not so much to do with relations like identity and survival but is a consequence of an ambiguity in our conception of what is valuable in life and what is lost by death.

Parfit's account of my relation to what we would normally regard as my own – and other people's – future selves suggests the possibility of thinking about our lives and deaths with a certain kind of detachment. For example, suppose that one of my friends is much influenced by my attitudes, beliefs, passions, interests and hobbies. Over time, my friend's interests flourish and develop while I undergo a radical change in personality, motivation and interests. At some suitably distant time in the future, there may be as much psychological connectedness between my present self and my friend's future self as between my present self and my own future self (see [5], §§94, 95). Notice that Parfit in effect adopts a no-ownership theory; we have to understand the locution '*my* present (future) self' and the like with this in mind. (What matters, for Parfit, is not the identity or ownership of selves but the psychological connectedness (involving memory, intention, ambition and so forth) between one set of experiences and other sets. Calling some of these experiences 'mine' is merely a *façon de parler*.)

In the imagined case, there will be a causal basis for the connectedness in question. My present self plays a role in causal processes, the outcome of which is the production of my future selves, even granted that these may be very different from the present one. But, equally, my present self has also played a role in causal processes which have resulted in the development of my friend's future selves. Seen in this light, the difference between myself and others may not be so great as we usually think it is. Not everything of value in my life may be lost when I die, for some future selves may still then exist and their projects, values and interests may have been the result of a causal process in which my present self plays a part.

But if we are happy to think of present and future selves in this way, we are immediately confronted with a problem, akin to the one discussed by Ehring. For why should my present self's *causal* role be a matter of any significance in the story? Think of the same kind of case, but without the causal ingredient. I do not value the existence of future selves living lives according to a certain conception of value, pursuing certain sorts of projects, motivated by certain sorts of desire simply because I (my present self, that is) happen(s) to play a causal role in bringing about such selves. Rather, if I try to play a causal role in bringing about the existence of such future selves (whether of mine or of others) it will be because I have certain values, care about certain sorts of projects and want to see certain kinds of desires fulfilled. But if this is so, my causal role cannot apparently be a significant part of the story. I should feel equally consoled, we might think, when facing the prospect of death, by the thought of my values, projects and interests being pursued whether or not I have any direct causal role in bringing about such a state of affairs.

We thus face a dilemma. If causal differences of the sort mentioned are trivial, then either personal survival is independent of causal role in the way just suggested or, if it does depend on causal role, it depends on merely trivial circumstances and is thus – given Parfit's view – no better off than personal identity. But if we abandon the causal condition on psychological connectedness this will allow – as Ehring recognizes – too many cases to count as cases of survival. The difficulty we face is that causal role seems to commend itself on the one hand, and yet to be irrelevant at the same time.

It is helpful for the diagnosis I am about to suggest if we get clear on two different kinds of causal contribution one item or event can make to the existence of some other item or event. These different contributions are associated with a further distinction between what I call *production processes* and *copying processes.* Consider a simple industrial process, the forming of bottle closures, say, from metal sheets. A stamping machine cuts discs from the sheets, and these discs are then pressed over a die to form the caps. The machinery for making such closures will produce hundreds and hundreds of extremely similar (in ordinary parlance 'identical') caps every hour so long as it is fed with similar sheets of unformed material. Any cap so produced will appear to be a replica of any other. Yet no one of the caps is copied from any other. The exact similarity of the caps produced is due to the uniform nature of the original unstamped sheets, together with the fact that all the caps are products of the same producing causes.

A copying process, by contrast, is one in which some particular item is used as an original from which other items are made. These other items have at least some of the characteristics they do because of their possession by the original. When I make a single copy of a printed sheet using a photocopier, the copier photographs my original sheet and then prints a new sheet as part of the larger copying process. Note that, when I decide to take more than one copy of the sheet, I initiate a production process: for although none of the resulting photocopied sheets are copies of each other, they are products of the same producing causes. It is perfectly normal for production processes to involve copying processes in this way.

It would be wrong to think that items can only make a causal contribution to the existence of subsequent items provided they are the originals for a copying process. Think for example, of a sewing machine. Once threaded and set to a certain stitch length, the machine will produce stitches of similar size. No stitch in a row of stitches is copied from any of the others, yet in a row of continuous stitches, each individual stitch makes some small contribution to the existence of the ones made subsequently simply by being a product in the production process: we could express this by saying that there are circumstances in which, if the machine had not produced a particular stitch successfully, then it would not have produced the following one. The important thing is that the causal contribution of one stitch to the existence of subsequent ones — however, minimal in this case — is not that of original to copy.

This distinction enables us now to identify different kinds of causal role that persons may have in various production and copying processes. Telecloning and teletransporting — much loved fictional processes for the creation of replica persons — are clear cases of copying processes, but where they churn out more than one replica, they are examples also of production processes. It is not surprising that the relation of each replica to the original is different from that of the replicas to each other. (Interestingly, Parfit gets into an unnecessary tangle about amnesia through a failure to appreciate this point. For an explanation of the problem and its solution, see [2].) More tricky than these fictional cases is the case of real life, however, where we find ourselves in a difficulty about deciding whether persons are the originals for copying processes, or mere ingredients in larger production processes.

We can start to get a feeling for why this is so if we reflect on two different conceptions of personal survival associated with two different — and much simplified — conceptions of what constitutes a person. Consider the case of parents surviving in their children. Even if such a notion of survival is metaphorical, it is, as I have argued elsewhere, important to the understanding of Parfit (see my argument in the final section of [1]).

At the risk of caricature, let me give brief details of two conceptions we might hold about such a case. On one account, we think of the gene pool as ultimately the producer of persons. We can take a fairly extreme variety of genetic determinism as our exemplar here, for it helps point the contrast to be made. On this view, what I pass on to my children is something distinctive of me only to the extent that I pass on some of the recipe responsible for many of my behavioural characteristics, including my choice of mate. I am one component in a larger production process, and what survives in my children, we might say, is not so much me as portions of the recipe that made me the way I am. By contrast, a quite different account of parental survival would emphasize that parents are the possessors of fine copying devices which ensure that some of their morphological and behavioural features are passed on to their children. By choosing to have children, we choose to use such devices, though recognizing that at present we have little control over them. At the same time, in our society, parents have an opportunity for developing certain kinds of characteristics and fostering certain interests and values by the way they bring up their children. Parents who, by building on the heritable basis, ensure that their children are as much like them as possible may survive to some significant degree in their children. (Of course, some parents, wary of conditioning their offspring and anxious to promote their autonomy, may try to ensure that they do *not* survive in their children.)

I am not in the business of defending either of the two views just canvassed. All that matters for the point I wish to make is that they be intelligible. On the latter view, parents play a causal role, whether conscious of it or not, in bringing about certain characteristics in their offspring, a role that is similar to that of the original in a copying process; to the extent that some of these characteristics are in the children because they were present in the parents, the latter survive in their children. Since successful copying is a matter of degree we can reflect (either with relief or regret) that in this case, high quality copying is impossible. This limits, of course, the degree to which a parent can survive in a child. On the first account, however, the situation is quite different. Parents instantiate, so to speak, genetic characteristics some of which are also instantiated by their children. Parents also play a causal role in bringing about the existence of their offspring, but this is not like the role played by the original in a copying process. The parent no more survives in the child than the enlarger in a photographic print. What survives in the child is merely some of the recipe that produced the parent.

How do these distinctions apply to the puzzle about causation and personal survival? Quite simply, there appears to be an ambiguity in our conception of what matters, as Parfit would put it, when we are thinking about persons. On the one hand, we value projects, pieces of behaviour, even gestures and grimaces, because of their association with people we value as individuals. On the other, we seem to value people in virtue of their projects, styles of behaviour and so forth. But this ambiguity infects our assessment of what is of interest in claims about survival. We are glad to see some of an old friend's wit, optimism and integrity in one of his or her children. If this is associated with other characteristics – a typical smile, a way of walking, or whatever, then we may take comfort that something of Smith survives in the child. But what? Is it that Smith, the individual person, somehow survives in that other person, or is it that characteristics and modes of behaviour that Smith happened to manifest have survived? Is it the behaviour and characteristics that we value, and is our valuation of Smith and Smith's offspring derivative from this? Or is it Smith the individual we value primarily, and the behaviour derivatively?

The distinction I am trying to make here is a bit like a kind of type/token one, but not quite. If we were to use the vocabulary, the ambiguity in our conception would be between taking persons as types and as tokens. To the extent that we cannot get clear on the distinction between persons as hardware and software, individual tokens and manifestations of types, or whatever, there is not only unclarity in the concept of a person, but room for alternative conceptual development. (See Parfit's discussion of types and tokens in [5], §99, and my remarks on developing concepts of the person in [3] chapter 10.)

The opposition, we think, cannot be simply this. In some way, we value both Smith and Smith's characteristic style, integrity and wit. But the 'in some way' simply fudges over the difficulty. Until we can sort out these different modes of valuing – of counting things (and people) important – we will not be able to come up with a principled means of dealing with the survival puzzle. Why this is so should now be clear. If we are interested in the survival of Smith, the individual, then our interest is in the sort of contribution Smith has made as an original in a process akin to a copying process. Smith's causal role is of necessity important to the question of survival, thought of in this way.

However, if our interest is primarily focused on style, modes of behaviour, integrity, wit and the like, our concern is likely to be that such valued things continue even after Smith's death. But Smith's own contribution to their continuation no longer seems so significant. They may survive because of Smith's role in a causal process productive of them; but they may survive anyway, in other persons who simply happen to be like Smith in the relevant respects. On this perspective, my own – or Smith's – efforts to cause future persons to adopt certain projects, styles of living and the rest may be worthwhile but, in the end, unimportant. Even the diminished causal role the genetic determinist allows us in the conception of persons as contributors to a production process for genes, no longer seems to carry much significance.

The simple answer in the face of the ambiguity indicated is to say that *personal* survival depends on causal role, in particular, on the sort of causal role an original has in a copying process. But it is not clear to me, nor, I suspect, are we likely to get much general agreement, that personal survival, as thus defined, captures what is important in the lives of persons. And so we face an impasse. We have progressed somewhat beyond the recognition of indeterminacy in judgments about personal identity and personal survival. But if my diagnosis is plausible we are caught between distinct modes of valuing persons and their behaviour which are bound to infect arguments about survival with a crippling ambiguity. We do not know quite how to express what is valuable in life, nor what is lost by death.

University of Stirling,
Stirling FK9 4LA

© ANDREW BRENNAN 1987

REFERENCES

[1] A. Brennan, 'Personal Identity and Personal Survival', *Analysis* 42.1 (January 1982), 44–50.
[2] A. Brennan, 'Amnesia and Psychological Continuity', *Canadian Journal of Philosophy*, Supplementary Volume II, 1985, 'New Essays in the Philosophy of Mind', pp. 195–209.
[3] A. Brennan, *Conditions of Identity*, Oxford, Clarendon Press, forthcoming 1987.
[4] D. Ehring, 'Survival and Trivial Facts', *Analysis* 47.1 (January 1987), 50–4.
[5] D. Parfit, *Reasons and Persons*, Oxford, Clarendon Press, 1984.

PARFIT'S ARGUMENTS AGAINST PARTIALLY RELATIVIZED THEORIES OF RATIONALITY

By RICHARD M. GALE

A COMMONLY held theory of rationality holds that it is rational for an agent to give primary importance to the reasons (desires, interests, needs, etc.)[1] that she has but not necessarily to those she has at the time of her action. To require the latter as well would do away with prudence, which hardly seems the rational thing to do. Derek Parfit, in his highly original and stimulating *Reasons and Persons*,[2] has presented arguments against such a partially relativized theory. This paper has the expressly negative purpose of rebutting these arguments.

---

[1] For the sake of brevity I will drop mention of the items within the parentheses.

[2] Clarendon Press: Oxford, 1984. All references to this book will be included in the body of this paper.

Parfit's arguments for the all-or-nothing theory of relativization are developed in the course of his defence of the following example of a fully relativized theory, called the 'Present-aim Theory':

> (P2) Reasons for acting *must* be fully relative. We should reject claims which imply that reasons can be *incompletely* relative. Thus we should reject the claim that reasons can be agent-relative but temporally neutral. We can call such claims incompatible with *full relativity.* (p. 141)

Notice that it would be redundant for a theory that relativizes reasons to those possessed by the agent at the time of her action[3] to do likewise for the reasons she has at the place of her action. Necessarily, the reasons she has at the time of her action coincide with those she has at the place of her action, since it is necessary that no more than one agent occupies a place at a time.

The first argument for (P2) charges the denier of it with being inconsistent. It is based on the requirement that parallel treatment be accorded to persons and times because there is a 'formal' analogy between the indexical terms, 'I' and 'now' respectively, that are used to denote them.

> ... the word 'I' refers to a particular person *in the same way in which* the word 'now' refers to a particular time. (p. 140)

Thus, the key premiss in Parfit's consistency-argument can be formulated as:

1. If any word 'w' refers *in the same way in which* the word 'I' does, consistency requires that any theory of rationality that relativizes reasons to a person do likewise for the reasons that the person has at the referent of her tokening of 'w' at the time of her action.

It then is pointed out that 'now' refers in the same way in which the word 'I' does, from which it follows that consistency requires that any theory of rationality that relativizes reasons to a person do likewise for the reasons that the person has at the referent of her tokening of 'now' at the time of her action.

It is imperative that we get clear about what Parfit must mean by the qualification 'in the same way' in 1. Since 'I' is indexical, is the similarity between 'w' and 'I' sufficiently tight so as to require parallel treatment by a theory of rationality if 'w' also is indexical?[4] This is too weak and would land Parfit in an inconsistency. 'This place' is indexical. And if reasons are to be relativized to those the agent has at the place she denotes by her tokening of 'this place' at the time of her action, it is possible that she violate (P2) – the *Present*-aim Theory – by giving primary importance to a reason she has earlier or later than the time of her action, the reason being that she occupies this ostended place (if she occupies it at all) at some time other than that at which she acts. Neither will it do for Parfit to restrict 'w' in 1 to indexical terms that are used without any accompanying act of ostension. For the indexical word 'formerly' is such a term. But again there is a clash with (P2), since she must give primary importance to a reason she has earlier the time of her action. Although 'formerly' is non-ostensive, it is relational, meaning 'earlier than now'. Thus, Parfit can restrict 'w' to indexical terms that are both non-ostensive and non-relational, to be called 'pure indexicals'.

---

[3] For the most part Parfit speaks about reasons 'at the time of acting' (pp. 140–4), but he also speaks about a reason that has 'force for me now' (p. 141) and 'what it is rational for me to do now' (p. 143). The latter could be interpreted to mean the reasons the agent has while deliberating, but for the sake of consistency they will be interpreted as referring to the time of the acting.

Why should the partial-relativizer accept this restricted version of 1? Parfit gives no explicit argument for 1 and thereby is open to the charge of begging the question against the partial-relativizer. The opponent of 1, however, can do more than just say 'Show me!'. First, a weak inductive argument can be advanced against 1, in which it is shown that there are concepts in which different pure indexicals are not implicated in analogous ways. For example, 'here' and 'now' enter into our concept of a material object in disanalogous ways, for whereas it is a conceptual truth that it is possible that a material object exist here both now and then, it is conceptually false that it is possible that a material object exist now both here and there. If it should be objected that an object can now occupy two different places by having different spatial parts occupy these places, it can be countered that it is not different temporal parts of an object that exists at two different times at the same place, since objects, as contrasted with their histories, do not have temporal parts. Further examples will be given below in which pure indexicals are not accorded parallel treatment in a conceptual anlysis. This gives us some reason, however slight, for doubting that the pure indexicals 'I' and 'now' should be accorded analogous treatment in our analysis of rationality. Not only is the epistemic scale tilted against 1; one possible way of supporting 1, viz. by showing that there is inductive evidence that in general pure indexicals figure analogously in concepts, is blocked.

There is available to the opponent of 1 a more powerful argument that shows grounds for believing that in respect to our very concept of rationality or rational agency 'now' and 'here' function disanalogously. There are axiological preferences of agents that show an asymmetry between past and future for which there are no analogous spatial preferences, in which the spatial (temporal) analogue to a given sentence is formed by replacing every temporal (spatial) term with a corresponding spatial (temporal) one, 'now' to be replaced by 'here', a date by a name of a place and 'later than' by any spatial relation term of your choice, say 'in front of' – it makes no difference. By showing that the pure indexicals 'here' and 'now' behave disanalogously in our concept of rational agency, it cuts the ground from under 1's requirement that 'I' and 'now' be accorded analogous treatment by an analysis of rationality solely on the basis of their being pure indexicals.

---

[4] An indexical word is used according to a referring rule that takes us from a context of tokening to a value of referent, the relevant contextual feature for 'I' being the tokener, 'now' the time of tokening, 'here' (when used without any accompanying act of ostension) the place of the tokener, etc.

At the core of all these temporally asymmetric preferences for which there are no spatial analogues is that whereas causation cannot go backwards, and thus it is conceptually true that

> T1. It is possible for an agent now to bring about intentionally (deliberate about, choose in respect to, etc.) an event later but not earlier than now,

it is conceptually false that

> S1. It is possible for an agent here to bring about intentionally (deliberate about, choose in respect to, etc.) an event in front of but not to the rear of here.

since causation can go in any spatial direction. An 'agent' is defined as one who deliberates and acts in accordance with her decisions in a morally responsible way. An agent is a conscious intervener into the on-going course of nature who attempts to bring about ends that she deems rationally desirable, the most important of which is her own full self-realization. Thus understood, it is a conceptual truth about an agent that

> T2. An agent holds that if it is better to be in state *Y* than state *X*, then it is better to be in state *X* now and state *Y* later than rather than earlier than now, provided that she is a free and active cause in bringing about state *Y* out of state *X*.

but not a conceptual truth that

> S2. An agent holds that if it is better to be in state *Y* than state *X*, then it is better to be in state *X* here and state *Y* in front of rather than to the rear of here, provided that she is a free and active cause in bringing about state *Y* out of state *X*.

provided that the states *X* and *Y* are described in a temporally pure way so that no demand is made on what obtains or fails to obtain earlier or later than X and Y.[5] In certain circumstances an agent might have a preference of the sort described in S2, but it would depend on extra-spatial information, whereas the agent's preference in T2 does not depend on extra-temporal information.

Another striking conceptual disanalogy between 'now' and 'here' is that whereas it is conceptually true that

T3. An agent regrets that her existence does not extend indefinitely later than now.

it is not conceptually true that

S3. An agent regrets that her existence does not extend indefinitely in front of here.

Agents want to be temporal fat men in the future direction but do not want to be spatial fat men in any direction. Since an agent has as her *summum bonum* her full self-realization and there is no upper limit on the possibility for the development of her character, abilities, and knowledge, it follows that it is natural and proper for an agent to regret her future demise since this limits the possibility for her full self-realization.[6]

Parfit has another formulation of his consistency-argument based upon a formal analogy between 'I'- and 'I-now'-sentences rather than between 'I' and 'now'. He writes:

> (P3) Consider any pair of claims that are related in the following way: the first claim contains the word 'I', but does not contain the word 'now'. The second claim is just like the first, except that it *does* contain the word 'now'. Call such a pair of claims *analogous*. If the first claim conflicts with the second, it is incompatible with full relativity, and should therefore be rejected. (p. 141)

The partial-relativizers response to (P3) can mirror her response to premiss 1. First, since (P3) is offered as a bald, unsupported claim, it is question-begging. Second, it does not hold true in general; e.g. it is true that I exist (tenselessly) at many different times but false (or maybe ill-formed) that I exist (tenselessly) at many different times now. These sentences are analogous in the sense specified in (P3) and conflict in that they differ in truth-value; but, *pace* (P3), it is the latter rather than the former sentence that is to be rejected. Furthermore, this very disanalogy gives positive reason for rejecting a version of (P3) that is restricted to sentences about rationality; since I exist at different times, it is rational in the sense of prudential for me now to consider in my deliberations what my future reasons will be and to give them at least equal weight with my present reasons.

---

[5] For an account of temporally pure descriptions see my book, *The Language of Time* (London: 1968), chapter 5.

[6] It might be, as Parfit claims on pp. 176–7, that we would be happier if we lacked the T3-type of bias toward the future and learned to take the passive, contemplative stance to the world of his character, Timeless, who is a non-agent. Even if this were true, it would be of no avail for protecting his (P2) theory of rationality from the above *agency*-based axiological disanalogies. Furthermore, since a theory of rationality is supposed to serve as a guide to *agents* in their deliberations, it is the moral intuitions of agents (not Timeless) that should be relevant to assessing a theory of rationality.

Parfit has another argument for (P2) that is not based on any formal analogy but instead on a special way in which a person is related to her own present mental states that is supposed to justify her having a bias in favour of her present reasons.

> My reasons for caring about what happens to me differ in kind from my reasons for caring about what happens to other people. The relation between me and my own feelings is more direct than the relations between me and the feelings of others.... In the same way, my reasons for caring about what is happening to me now differ in kind from my reasons for caring now about what did or will happen to me. The relation between me now and what I am feeling now is more direct than the relations between me now and my feelings at other times. (pp. 145–6)

The crucial question is what does 'more direct' mean? Parfit does not tell us, and thus it is up to us to search around for some interpretation that can be used to support (P2).

The most natural way to read it is as signifying degree of certainty regarding the feelings (reasons, etc.) she has at that time. On this interpretation Parfit's theory of rationality would have to be revised so that the degree of importance an agent is to assign to a reason is to be correlated with her degree of certainty regarding its existence. But this fails to support (P2); quite to the contrary since an agent often is just as certain about one of her former reasons, as well as about the reasons of others, as she is about her own present reasons. Reasons, unlike feelings (e.g. pains), are not the sort of things to which one has privileged access and about which one can have incorrigible knowledge.

There is another problem posed by this interpretation for Parfit's claim that an agent should give primary importance to the reasons she has *at the time of action.* A theory of rationality should serve as a guide to deliberation. But deliberation precedes action. And, thus, the Parfitian deliberator must try to infer what reasons she will have in the future when she carries out her decision. But such inferences, in many cases, do not yield as high a degree of certainty as she has for the reasons she possesses while deliberating. And, thus, *pace* (P2), she must give primary importance (or a greater degree of importance) to a reason she has prior to the time of her action.

Another possible reading of 'more direct' is that of having exclusive mental ownership. At a given time the agent and only the agent has such a cartesian ownership of her reasons, desires, etc. This is dubious, since having a reason, etc. is not a mental process, e.g. a feeling; reasons are shareable. Further, even if it were true that there is exclusive mental ownership of reasons, it still is true that an agent will be the exclusive cartesian owner of her future reasons. And this gives her good reason to give as much importance to these future reasons as she accords to her present ones.[7]

University of Pittsburgh,
Pittsburgh, Pennsylvania 15260,
U.S.A.

© Richard M. Gale 1987

---

[7] I profited from the criticisms and suggestions of Kurt Baier, David Carey, David Gauthier, Paul Hurley, and Shelly Kagan. I am especially indebted to the referee for *Analysis* whose searching criticisms caused a radical revision of this paper.

CHILDREN OF A LESSER GOD? A REPLY TO BURKE AND CRISP

By P. J. McGrath

MICHAEL B. BURKE ('Theodicy with a God of Limited Power; a Reply to McGrath', Analysis 47.1, January 1987, pp. 57–8) and Roger Crisp ('The Avoidance of the Problem of Evil: a Reply to McGrath', Analysis 46.4, October 1986, p. 160) appear to have missed the point of my article 'Evil and the Existence of a Finite God' (Analysis 46.1, January 1986, pp. 63–4). What I was trying to establish was not that the problem of evil remains (or becomes) insoluble when God is conceived as limited either in goodness or in power, but the more modest claim that the problem is not thereby eliminated. The fact that Burke and Crisp find it necessary to suggest ways in which the presence of certain types of evil in the world could be made compatible with the existence of a benevolent, though finite, God serves to reinforce my point. However, since my arguments could be used in support of the stronger claim, I shall consider their solutions briefly to see how far, if at all, they are successful.

Burke's answer to the question: why would a powerful but non-omnipotent God not have eliminated evils such as smallpox? is as follows:

> McGrath overlooks the reply that God failed to eliminate smallpox not because such a feat was beyond his power, but because all of his vast but finite power was required for the elimination or prevention of *greater* evils. It could be suggested, to give just one example, that our galaxy is but one of hundreds of millions with intelligent life, that God can attend to 'only' a dozen galaxies at a time, and that ours is not among those most urgently in need of attention.

This is no doubt a possible explanation for God's failure to prevent or eliminate certain evils. But has it any claim to credence? One rather obvious difficulty with it is that it offends against the principle of simplicity. To account for the divine neglect Burke has to postulate myriads of planets containing intelligent beings, many of them afflicted with even greater evils than those found on earth. It could perhaps be urged that there are independent reasons for thinking that the universe contains other communities of intelligent beings. But these reasons are suggestive rather than conclusive and the supposition that the evils which exist on earth are surpassed by those found elsewhere is little better than a conjecture. Even if we were to concede the point that the reasons for thinking that there are extra-terrestrial communities of intelligent beings also constitute reasons for thinking that they are afflicted with similar evils to those found on earth, this, as we shall see in a moment, would seriously weaken the explanatory value of Burke's hypothesis. In any event, its lack of strong evidential support makes it impossible to avoid the conclusion that it should be rejected in favour of a simpler explanation – such as that God does not exist or, if he does, that he is indifferent to human welfare.

The second difficulty with Burke's hypothesis is that it appears incapable of doing the job which he assigns to it. The evils under discussion are ones which have been, or very likely will be, eliminated through the advancement of science. All that is required on God's part therefore is that he anticipate human progress by communicating the relevant information to man so that as much evil as possible be eliminated without delay. (This task would be enormously simplified if the evils that exist on earth are similar to those found elsewhere; on the other hand, the claim that there are evils elsewhere which are dissimilar to those found on earth is mere supposition.) To suppose that God does not possess such knowledge or that he is incapable of communicating it successfully is to reduce his status to that of just another intelligent being; there would no longer be any good reason for treating him as divine or as a proper object of worship.

The same objections may be urged against Crisp's hypothesis that 'there are two powers at work in the universe, one infinitely good, the other infinitely evil'. Since this appears totally devoid of independent evidential support, it too should be rejected in favour of a simpler explanation. Likewise it appears incapable of explaining why God does not eliminate certain evils by supplying the relevant information to man.

It is of course possible that Burke and Crisp intended to do no more than show that the existence of a benevolent God is logically compatible with the continued presence in the world of evils such as smallpox. This would save their explanations from the threat of Ockham's razor, but at the cost of depriving them of any bearing on the question: does the existence of certain evils not provide compelling evidence against the existence of a benevolent God?

Finally, I am surprised that Crisp thinks that I failed to notice 'the so-called "Acts and Omissions Doctrine"... [that] God is not as responsible for his omissions as for his acts'. I did not refer to this doctrine in my article since it was unnecessary to do so if I was right in arguing that 'a God who tolerated the sort of evils which actually exist simply because of his lack of concern for the welfare of others would be a moral monster'.

University College, Cork, Ireland © P. J. McGRATH 1987

REPLY TO WREEN

By JENNIFER HORNSBY

A PERSON may set out to contract particular muscles, and succeed in doing so when he clenches his fist. Cases of this sort are familiar, and they have cropped up in the philosophy of action for a variety of reasons. In 'Two "by" ways' (ANALYSIS 47.2, March 1987), Michael Wreen introduces a problem by quoting a fuller description I gave of the particular case. The problem is that that description (quoted by Wreen), though it comes naturally, is absurd.

Wreen introduces what is I think another problem by reference to the same case. As he put it (p. 121):

> The situation described is basically one in which we are tempted to subscribe to both:
>
> (1) John clenched his fist by contracting his muscles
>
> and
>
> (2) John contracted his muscles by clenching his fist.

One can see that this is a different problem, because it does not arise unless one accepts Wreen's (2) and certain assumptions about 'by'; whereas (2) was not used in the description whose absurdity constitutes the problem I had presented, and the particular absurdity was present without any assumptions about 'by'.

When Wreen says that 'Hornsby's solution to the puzzle won't do', he seems to conflate the two problems. He takes my attempt to solve the first as competing with his attempt to solve the second. But nothing he says can count against my claim that, in order to see our way through the first problem, the denotations of transitive and intransitive verbs must be distinguished. And if we are coherently to say everything we want to about the case, then we have to see our way through both problems.

(Even if Wreen is not interested in the first problem, like Wreen, I have been interested in the second. Wreen's quotation was from *Actions*, Routledge and Kegan Paul, London, 1980, pp. 20–1. My account of the second problem is also there: it is drawn together at p. 93, and it contains arguments against the view of that problem which Wreen presents in his paper. At p. 28, and again at p. 94, I went into more detail than I have here about why the two problems are different.)

Corpus Christi College,
Oxford OX1 4JF

© JENNIFER HORNSBY 1987

'"BY" AND "BY"'

By ALAN R. WHITE

IT is, I believe, (1) physiologically true that one clenches one's fist by contracting certain muscles, and (2) performatively true that one contracts certain muscles by clenching one's fist. We are reluctant to accept that (2), like (1), gives a causal explanation, since it would be backward causation. Miss Jennifer Hornsby (*Actions*, 1980, ch. II) sought to avoid this conclusion by appealing to the difference between the transitive and the intransitive uses of 'clench' and 'contract'. Mr Michael Wreen, rightly in my opinion, recently argued against this ('Two "by" ways', ANALYSIS 47.2, March 1987, pp. 120–4). He suggested, alternatively, that in (1) 'by' indicates causation, but that in (2) 'by' means 'and this was indirectly achieved or performed by' analogously to the way in which, for example, one *a fortiori* proves something about all natural numbers by proving it about all real numbers. Though I think that he is correct that the solution to the problem lies in the role of 'by' in the two sentences, I don't think he is right that it plays an ambiguous role.

I suggest a different solution. First, take some more analogies. It is, I believe, (1) chemically true that one boils or freezes a liquid by raising or lowering its temperature and, further, that one raises or lowers its temperature by increasing or decreasing the kinetic energy of its molecules, and (2) performatively true that one increases or decreases the kinetic energy of the molecules of a liquid by raising or lowering its temperature and, further, that one raises or lowers its temperature by boiling or freezing it. It is not, I believe, chemically acceptable to say that boiling and freezing cause changes in temperature or that changes in temperature cause changes in kinetic energy or to say that an alteration in kinetic energy is due to a change in temperature or a change in temperature due to the boiling or freezing; but only the converse of these. Nor is the problem solved by appealing to the undoubted transitive and intransitive uses of 'boil' and 'freeze'. We could, indeed, eliminate the transitive use by phrasing the problem as 'The kinetic energy of the molecules of a liquid alters by the liquid's boiling or freezing' and 'A liquid boils or freezes by an alteration in the kinetic energy of its molecules'. The conceptual problem of the role of 'by' here is not related to the notion of action at all. Nor is the problem solved by suggesting that by boiling or freezing a liquid we 'indirectly or *a fortiori* achieve or perform' the task of altering the kinetic energy of the molecules.

The solution, I suggest, is this. To say that one does one thing, *B*, *by* doing another, *A*, is unambiguously to indicate that *A* is a *sufficient condition*, whether it be causal, logical, or of some other kind, of *B*. Thus, as in (1), one clenches one's fist by contracting certain muscles, boils a liquid by raising its temperature, raises its temperature by increasing the kinetic energy of its molecules, sweats by opening the pores in one's skin, kills someone by stopping his heart, says something true by stating a fact, draws a triangle by drawing an equilateral triangle, says something true by saying something known, turns on the light by flicking a switch, signals by waving, does one's duty by staying at one's post, gains entrance to a building by forcing a window, passes through Leicester by taking the Nottingham to London train, etc.

If, in addition, *A* is a *necessary condition*, whether it be causal, logical, or of some other kind, of *B*, then it follows logically that *B* is a *sufficient condition* of *A*. Hence, one can say in one and the same sense of 'by' that, in such cases as (2), one does *A by* doing *B*. It is because the *A*'s in contracting certain muscles, raising a liquid's temperature, increasing the kinetic energy of its molecules, opening skin pores, stopping the heart and stating a fact are all also necessary conditions for their corresponding *B*'s, while the *A*'s in drawing an equilateral triangle, saying something known, flicking a switch, forcing a window, waving, staying at one's post, and taking the Nottingham to London train are not necessary conditions for their corresponding *B*'s, that the *B*'s of the former, but not of the latter, are sufficient conditions for their *A*'s and, hence, can be preceded by 'by'.

If *A* is logically sufficient for *B*, then *B* can be, but need not be, logically sufficient for *A*; that is, *A* and *B* can mutually imply each other. But if *A* is causally sufficient for *B*, *B* cannot be causally sufficient for *A*; that is, *A* and *B* cannot mutually cause each other. Which of the two is the cause and which the effect is an empirical, not a conceptual, question. But *A* cannot be both the cause and the effect of *B*, despite some traditional talk of something's being *causa sui*.

Equally, if *A* is sufficient for *B* because in such and such circumstances *A* amounts to *B*, for example by waving one signals, by staying at one's post one does one's duty, then *B* cannot, at the same time, be sufficient for *A*, for example one does not wave by signalling or stay at one's post by doing one's duty.[1]

The University, Hull

© ALAN R. WHITE 1987

---

[1] I am indebted to my colleague, Paul Gilbert, for corrections of an earlier draft.

the review of

metaphysics

a philosophical quarterly

MARCH 1987 | **VOL. XL, No. 3** | **ISSUE No. 159** | **$8.00**

articles

ROBERT SOKOLOWSKI	*Exorcising Concepts*
ARTHUR RIPSTEIN	*Explanation and Empathy*
KENNETH G. FERGUSON	*Truth Conditions for "Might" Counterfactuals*
ROBERT E. WOOD	*Image, Structure and Content: On a Passage in Plato's Republic*
THOMAS J. BOLE, III	*Contradiction in Hegel's Science of Logic*

critical studies

ROBERT B. PIPPIN	*Blumenberg and the Modernity Problem*
JANICE FLOERSCH and Staff	*Summaries and Comments*

philosophical abstracts *announcements*

Individual Subscriptions $21.00 Institutional Subscription $32.00 Student Subscriptions $1

Jude P. Dougherty, Editor
The Catholic University of America, Washington, D.C. 20064

Philosophical Studies

An International Journal for Philosophy in the Analytic Tradition

Editor-in-Chief
John Pollock, *Dept. of Philosophy, University of Arizona, Tucson, USA*

Philosophical Studies is devoted to the quick publication of analytical contributions, particularly (but not exclusively) in epistemology, philosophical logic, the philosophy of language, and ethics. The principal aim is to publish articles that are models of clarity and precision in dealing with some significant philosophical issues. Articles in the journal are intelligible to philosophers whose expertise lies outside the subject matter of the article.

Subscription Information **ISSN** 0031–8116
1987, Volume 51–52 (6 issues)
Institutional rate: Dfl. 412.00/US$172.00 incl. postage/handling
Private rate: Dfl. 156.00/US$65.00 incl. postage/handling

Private subscriptions should be sent direct to the publishers

D. Reidel Publishing Company

Kluwer academic publishers group

P.O. Box 989
3300 AZ Dordrecht
The Netherlands

101 Philip Drive
Norwell,
MA 02061 U.S.A.

Falcon House
Queen Square
Lancaster LA1 1RN U.K.

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol. 64, No. 3; September 1986

EDITOR: BRIAN ELLIS

ARTICLES

Richard Warner	A Challenge to Physicalism
Robert Heinaman	Incompatibilism Without The Principle of Alternative Possibilities
Michael McDermott	Narrow Content
Alec Hyslop	Emotions and Fictional Characters
John Wright	Field and McDowell on Reference
Robert Kirk	Sentience, Causation and Some Robots
Robert P. Amico	On the Vindication of Deduction and Induction
F.C. White	On a Proposed Refutation of Relativism

DISCUSSION

F. John Clendinnen	Induction, Indifference and Guessing
Reed Richter	Further Comments on Decision Instability

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTIONS: Individuals A$17 (US$18 or Stg£10); Institutions A$34, (US$36 or Stg£18). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

OXFORD

Mind *The Journal of the Mind Association*

Editor: **Simon Blackburn**, Pembroke College, Oxford

Mind is the best known British philosophical journal, and one of the most widely read and respected journals in professional philosophy.

Forthcoming and recent papers include *Paul Churchland* on reductive strategies in cognitive neurophysiology, *Ruth Garrett Millikan* on the metaphysical nature of Putnam's anti-realism, *Jonathan Westphal* on colour, *Michael Pendlebury* on force in semantics and *Ian Hacking* on inverse arguments in cosmology.

ORDER FORM

Payment may be made by cheque or major credit card.

Please send me:

☐ *Mind* Volume 96 UK £16 (Individuals £10) Overseas £19 (£13.50)

☐ a sample copy

Name ....

Address ....

Postcode .... Country ....

Oxford University Press, Walton Street, Oxford OX2 6DP, UK

JOURNALS

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, T. R. Baldwin; T. S. Champlin, M. Clark, J. P. Dancy, Andrew Harrison, B. J. Heal, R. F. Holland, C. A. Kirwan, D. H. Mellor, H. W. Noonan, M. A. Stewart, A. R. White, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. ANALYSIS is published by Basil Blackwell Ltd, 108 Cowley Road, Oxford OX4 1JF, England and P.O. Box 1320, Murray Hill Station, NY 10156, U.S.A. Each volume comprises four numbers, each of at least 48 pages, appearing in January, March, June and October. Orders should be sent to the publisher. Subscriptions prices for 1986 are as follows:

	U.K.	Overseas (incl. Eire)	U.S.A. & Canada
Institutions	£12.00	£15.00	U.S. $33.00
Individuals	£8.00	£15.00	U.S. $21.00

BACK ISSUES. Volumes 1–37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

MICROFILM copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30–32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

ADVERTISING. For details contact Bill Drake, telephone (0869) 38477, or write c/o the publishers.

CONTRIBUTIONS. Articles submitted for publication should be addressed to the incoming Editor, Mr Peter Smith, Department of Philosophy, University College of Wales, Aberystwyth, Dyfed SY23 3DY.

Contributions are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable